LOGOS ARCHITEKTON

Journal of Logic and Philosophy of Science

ISSN: 2065-0469

Volum editat cu sprijinul CNCSIS, grant nr. 423/2007, cu tema *Paradoxurile implicației stricte*.

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruța Săcelean
Str. Hasdeu nr. 51
400371 Cluj-Napoca, România
Tel./Fax: (+40)-264-597.401
E-mail: editura@editura.ubbcluj.ro
http://www.editura.ubbcluj.ro/

LOGOS ARCHITEKTON

Journal of Logic and Philosophy of Science

Vol. 3 No. 1 Spring – Summer 2009

Realism /Antirealism

&

Vol. 3 No. 2 Autumn – Winter 2009

Miscellanea

Edited by Virgil Drăghici

CLUJ UNIVERSITY PRESS
2009

SUMMARY

REALISM / ANTIREALISM

MISCELLANEA

REALISM / ANTIREALISM

The Paradox of Consciousness and the Realism/Anti-Realism Debate

Eric DIETRICH & Julietta ROSE [1]
Binghamton University, New York

Abstract:

Beginning with the paradoxes of zombie twins, we present an argument that dualism is both true and false. We show that avoiding this contradiction is impossible. Our diagnosis is that consciousness itself engenders this contradiction by producing contradictory points of view. This result has a large effect on the realism/anti-realism debate, namely, it suggests that this debate is intractable, and furthermore, it explains why this debate is intractable. We close with some comments on what our results mean for metaphysics and philosophy, in general.

Keywords: Consciousness, supervenience, dualism, materialism, physicalism, realism, anti-realism, zombies, zombie twins, dialetheism, true contradictions, metaphysics.

1. Introduction

It is not often noted how paradoxical consciousness is. Even when philosophers explicitly discuss some paradoxical aspect of it, they usually view that aspect as a solvable problem rather than as something intrinsic to

---

1 There is no first author; the authors' names are alphabetized. Also, we thank Graham Priest for comments on an earlier draft of this paper.
E-mail: dietrich@binghamton.edu.

consciousness (e.g., Chalmers', "The paradox of phenomenal judgment" (1996, ch. 5)). This paper is about consciousness's paradoxical nature and its role in the realism/anti-realism debate. Since zombies are a natural and easy introduction to this paradoxical nature, we begin with them, using them to argue that dualism is both true and false. Then we widen our scope, locating the source of this paradox in the contradictory combination of points of view created by consciousness itself. We then argue that the paradoxical nature of consciousness is in turn responsible for one important strand of debate between realists and anti-realists. We close with some comments on what our conclusions mean for that debate, for metaphysics, and for philosophy, in general.

Two preliminary matters. First, we use a notion of supervenience to define dualism and materialism. But standard supervenience won't do the job required (see Horgan, 1993). We therefore use a version of Horgan's notion of *superdupervenience* which is defined as "ontological supervenience that is robustly explainable in a materialistically explainable way" (Horgan, 1993). We define superdupervenience thusly:

> *A* facts *superdupervene* on *B* facts iff any two possible situations identical in their *B* facts are *eo ipso* identical in their *A* facts, and the *A* facts are robustly explainable in terms of the *B* facts because of the "*eo ipso*" condition.

This definition differs from ordinary supervenience (*A* facts *supervene* on *B* facts iff any two possible situations identical in their *B* facts are identical in their *A* facts) in 1) the "*eo ipso*" condition, and 2) the epistemic contact between the two levels (which is from Horgan).[2] Superdupervenience guarantees that if X logically superdupervenes on the physical, then X is itself physical and explainable as such. Supervenience alone, even logical supervenience, doesn't secure this tight connection. Hence, superdupervenience implies supervenience, but not vice versa. Now,

---

2 Our notion of superdupervenience appears to be somewhat stronger than Horgan's.

we define "dualism" as the thesis that consciousness doesn't logically superdupervene on the physical (see the appendix). Briefly, fixing all the physical states of the universe is not sufficient to fix (or guarantee) the phenomenal states in the universe. Materialism (or physicalism, we shan't distinguish the two), then, is the thesis that consciousness *does* logically superdupervene on the physical.

Secondly, we take dialetheism seriously. Dialetheism is the claim that some contradictions are true (they're false, as well, but also true). Not all contradictions are true, of course, and *a fortiori* not all statements are true. That is, *ex contradictione sequitur quodlibet* ("from contradiction, everything follows") is false. Dialetheism is well-defended by Priest (2006). It is used to great advantage in Priest (2003). Dietrich (2008) presents an intuitive, easy to follow way to see that a certain contradiction is true.

Paper map: In section 2, we introduce the central problem with zombie twins when used to argue for dualism. In section 3, we present an argument based on this problem showing that dualism is both true and false. The best way out of this contradiction is to reject zombie twins as impossible, a move which has a lot to recommend it. However, in section 4, we show that zombie twins are possible if dualism is true, and we argue that there are good reasons, independent of zombie twins, to think that dualism is, in fact, true. In section 5, we show how our analysis of the zombie issue extends to the realism/anti-realism debate. Specifically, we show that this debate is unresolvable, and that there are good reasons for thinking that both anti-realism and realism are true. We then close with a comment about what our results mean for philosophy in general.

2. Do Zombies Dream of Zombie Twins

Besides being undead, unconscious, and unnerving, zombies also create logical problems. If it weren't for this last property, the first three would probably be tolerable. That zombies, specifically, zombie twins, cause logical problems is well-known (see, e.g. Chalmers, 1996, Dennett, 1995, Moody, 1994). What is less appreciated, at least by some (e.g., Flanagan and Polger, 1995) is how deep these problems run.

The difficulties with mere zombies (unconscious creatures merely resembling humans in one way or another, e.g., functionally or behaviorally) versus those with zombie twins (unconscious creatures physically identical to us) are not equal in virulence; zombie twins are far more problematic. We focus on zombie twins.

The logical problem with zombie twins we will focus on has been called *conscious inessentialism* (Flanagan, 1992[3]). A central intuition had by those of us who are conscious (and have thought about it) is that being conscious is why we make the experiential judgments that we do. We believe "that looks red," "that tastes salty," "that hurts," "that feels good" because we consciously experience a red color, a salty taste, a pain, or a pleasure. We call this *conscious essentialism*: consciousness is essential to our mental lives having the contents they do (and not just phenomenal contents, but semantic contents as well).[4]

Conscious essentialism appears not only true, but obviously true. Yet, when using zombie twins to argue for dualism, this intuition has to go (Chalmers famously uses such an argument, 1996). The argument requires zombie twins to make the very same judgments we do. So, being conscious cannot be the source of such judgments. Instead, we are left with the unpalatable position that we who are conscious judge that an apple is red not because we experience its red color (i.e., not because it looks red to us), but solely because of the physical processes of our cognitive and perceptual systems. Zombie twins might establish dualism, but the cost appears to be rendering consciousness useless in our mental lives. Hence, the specific form of dualism established is something akin to epiphenomenalism or parallelism, neither or which are plausible.

This problem is good news to materialists (or physicalists) of various stripes, that is, those materialists for whom giving up the intuition that our

---

3 The term should probably be "conscious*ness* inessentialism" since it is a thesis about consciousness, but "conscious inessentialism" is already established, so we will use it and its related variants.

4 For us here, *conscious essentialism* is equivalent to *not* (*conscious inessentialism*).

conscious experiences *inform* us is completely out of the question. Such philosophers then follow this to the conclusion that zombie twins are not possible while admitting that they are conceivable in a rough or superficial sense. Other materialists insist that zombie twins aren't even conceivable, provided that the term "conceivable" picks out any sort of psychologically plausible type of conceiving (Dennett, 1995). Finally, this problem is also good news to some interactionists – dualists who think that the phenomenal realm crucially interacts with the physical realm to produce the conscious thoughts and concepts that we have on a daily basis. These interactionists embrace conscious essentialism. (It's because of this kind of dualism that we need superdupervenience: it prevents this kind of dualism from turning into materialism.)

One can view the work of philosophers who have been prepared to use zombie twins in arguments for dualism in terms of a cost-benefit ratio. Yes, zombie twins are expensive, but they are worth it, for they give us that which is most sought after by theorists of all stripes: a true, but shocking theory that upsets the apple cart of science.[5] In chapter five of his book, *The Conscious Mind*, Chalmers argues that paying the cost of using zombie twins yields unexpected epistemological and metaphysical rewards that will deeply inform a science of dualistic consciousness, that is, a science that takes dualism seriously.

But zombie twins are not merely problematic. In the next section, we analyze an argument that shows that dualism is both true and false. This argument rests on premises that are all arguably plausible, so it isn't obvious which of them should be abandoned, assuming any should. It is, though,

---

5 In an argument for dualism relying on zombies, zombie twins are required; mere zombies won't do. This is because dualism is the claim that consciousness is not a material property of minds *in our world* -- the actual world. To prove this requires producing a world physically identical ours but without consciousness. A world physically identical to ours would have to have physical replicas of *us* in it. Those creatures are our zombie twins. Of course, producing such a world is question-begging in this context, since such a move assumes that consciousness can be sundered from the physical, which is precisely what is at issue.

obvious which ones various disputants in the zombie debates will abandon; the problem is, their arguments run afoul of either conscious inessentialism or they make claims that are obviously false.[6] The paradoxical nature of our zombie twin argument runs deep, for even if, like us, one takes conscious essentialism to be non-negotiable and therefore concludes that zombie twins are impossible, one can still, on very plausible premises, conclude that zombie twins *have to be* possible. This we show in section 4. Then in section 5, we argue that the real problem traces through the conscious essentialism/inessentialism debate, back to consciousness itself, which in turn funds a central and intractable version of the realism/anti-realism debate.

3. The Contradictory Argument

Here is the argument that dualism is contradictory. Where not controversial, the justification is placed in square brackets. The controversial premises are: 1, 3, 4, and 7. We discuss them in section 3.1.

1. If some conscious agent conceives of its zombie twin
then dualism is true.
2. If humans in the actual world conceive of their zombie twins then so do their zombie twins (i.e., our zombie twins conceive of their zombie twins). [*Definition of "twins"*.]
3. For all *X*, if *X* conceives of its zombie twin then *X* is conscious.
4. Humans in the actual world conceive of their zombie twins.
5. Zombie twins conceive of their zombie twins. [*4, 2 and modus ponens*]
6. So zombie twins are conscious. [*5, 3, the relevant instantiation, and MP*]
7. If zombie twins are conscious, dualism is false. {*Because consciousness is revealed as a physical property. This means we've misconceived zombie twins, see below.*}
8. Dualism is false. [*6, 7, MP*]

---

6 Chalmers, for example, has to embrace conscious inessentialism, at least in some form. Dennett, for example, denies that we have qualia, i.e., conscious experiences (1988).

9. Dualism is true [*4, 1, the relevant instantiation, and MP; note: 4 and 3 give that humans are conscious*].

3.1. Two Paths Through the Contradictory Argument

There are two paths through this argument. One path – steps 4, 1, and 9 – assumes conscious inessentialism and the other – steps 4, 2, 5, 3, 7, 8 – assumes conscious essentialism. Both essentialism and inessentialism have strong pulls on almost all philosophers' thoughts about consciousness. The pull of essentialism is obvious: How could a person blind from birth know what it is like to see red? How could such a person have the appropriate phenomenal concept of red? The answers to both are "She couldn't." The pull of inessentialism can be seen via noting that you, the reader, might be the only conscious being in the universe. For all you know (in a very strong sense of "know"), everyone else in the universe might be a zombie, doing what they are doing totally bereft of consciousness. They talk about seeing red and the like simply because they picked up such locutions from you, not because they actually see red – they're zombies after all. The zombies that surround you are much like parrots who mimic human speech patterns but who don't actually know what they are talking about.[7] In sum, the pull of essentialism is strongest when we think about ourselves, and the pull of inessentialism is strongest when we think about other people.

What funds the Contradictory Argument is, therefore, contradictory assumptions about the role of consciousness in our mental lives. The argument is ambiguous between embracing essentialism and embracing inessentialism. In turn, this ambiguity reveals itself in the kinds of conceiving relevant to conceiving of zombie twins. The two paths differ also in the strength of their commitment to zombie twins. The conscious

---

7 See Valdman, 1997, for an excellent analogy between zombies and parrots who happen to live on a certain island that was home to a couple of castaway quantum physicists. The parrots talk all day about quantum mechanics and even stumble over new theorems, but of course don't know what they are talking about. See Moody, 1994, who argues that zombies could "talk" about red only if they were among conscious beings who also talked about red – zombies couldn't originate talk of red things. And for some cool stuff on parrots, see Pepperberg, 2000.

inessentialism path is strongly committed to the notion of zombie twins; the conscious essentialism path is only weakly committed, and in fact concludes that the notion is flawed in crucial ways.

3.2. The Contradictory Argument in Detail

Here are the justifications for premises 1, 3, 4, and 7.

Premise 1. If some conscious agent conceives of its zombie twin then dualism is true.

According to Chalmers, there are three requirements for this premise: 1) we must be able to conceive of zombie twins in the right way, 2) conceivability must imply possibility, and 3) the possibility of zombie twins must entail dualism (1996, 2002). The third requirement is guaranteed, according to Chalmers, by introducing the notion of *logical supervenience* (1996; we think logical superdupervenience is required, this change is easy to make). Chalmers claims that the first two can be achieved using his notions of *ideal, primary, positive conceivability*, and *primary possibility*, because the primary possibility of a given proposition (statement) is entailed by that proposition's ideal primary positive conceivability (Chalmers, 2002). The development of 1), 2), and 3) is in the appendix (which can be skipped, if the reader is content to just grant premise 1 or is already familiar with Chalmers's theories); here, we assume that these three requirements are met (the appendix demonstrates the reasonableness of this assumption). But there's an untoward consequence that one also must embrace if one is to accept premise 1: If zombie twins are possible, which premise 1 purports to show, then conscious inessentialism is true. For, our zombie twins think, do, and say exactly what we do. Since they aren't conscious, consciousness must be inessential to what *we* think, do, and say.

Premise 3. For all X, if X conceives of its zombie twin then X is conscious.

The argument for premise 3 is simply that our being conscious seems necessary for conceiving of our zombie twins. Conceiving of doing without something – anything – requires first having that thing, or at least

conceiving having it. Consciousness, however, is such that it seems quite unlikely that we'd conceive having it if we didn't (i.e., if we weren't conscious); we actually have to have it to conceive of not having it. So, zombie twins can't conceive of their zombie twins, as such. So if something does so conceive of its twin, it must be conscious and it's twin not. The etiology of our zombie twin intuition (the intuition that we each have one) remains far from clear, but zombie twins only make sense in a world with conscious beings in it to begin with, indeed, the very beings conceiving of their zombie twins have to be conscious.

Denying premise 3 is very expensive. (Chalmers denies this premise, see below. He asserts that zombie twins conceive of their zombie twins yet are not conscious (1996, ch. 5)). To deny this premise requires embracing conscious inessentialism. This in turn means that our zombie twins will produce arguments for dualism even though they are not conscious at all. There is nothing it is like to be a zombie twin, yet there they are arguing about inverted spectra and whether or not consciousness is a nonphysical property of the universe. And all this even though everything about zombies is physical – in the zombie world, everything logically superdupervenes on the physical. So being conscious is irrelevant to theorizing about consciousness, indeed, it is irrelevant to even having the intuition that we each have zombie twins (and clearly, some humans have this intuition, so their zombie twins must, too). All this is stunningly implausible (see section 4.1 below). But it must be embraced to deny premise 3.

There is a further complication with premise 3. Steps 5 and 3 together entail that zombie twins are conscious (step 6). But this seems to contradict the definition of zombie twins. One might think, therefore, that step 6 is contradictory: zombie twins can't be conscious. Hence, any argument that zombie twins are conscious must be fatally flawed. This is an important point. As we discussed above, premise 1 is only used in one path through the Contradictory Argument – the conscious inessentialist path. This path doesn't use step 6 at all, which is part of the separate, essentialist path through the Contradictory Argument. At root, what the essentialist path does is recognize that the concept of a zombie twin must be redefined (or, that the notion is incoherent). Thus: zombies are either not conscious and

hence behaviorally different from us (since consciousness is essential to our behavior), and hence they are *not twins*, or they are conscious and behaviorally the same as us, hence they are *not zombies*. Of course, a conscious essentialist could just assert that zombie twins are impossible because the notion is incoherent. Such a conscious essentialist would *not* have to be a materialist, she could be a dualistic interactionist of a certain sort.

Premise 4. Humans in the actual world conceive of their zombie twins.

This premise is clearly true, for a standard, superficial notion of conceiving, which is just bringing before the mind some appropriate referring expression. Anyone following this paper so far has conceived of her or his zombie twin in this sense. The question is, however, can humans conceive of their zombie twins in the *right way*, which uses ideal, primary, positive conceivability (see the appendix)? That this can be done is far less clear. The right kind of conceiving can be achieved, however, if one explicitly embraces conscious inessentialism. This can be accomplished by embracing, say, *parallelism*, the view that the physical realm and the phenomenal realm don't interact at all, but merely parallel one another (parallelism is also known as "pre-established harmony," which is the view of the situation touted by Leibniz). Once this is done, robustly conceiving of zombie twins using ideal, primary, positive conceivability is readily accomplished.

Premise 7. If zombie twins are conscious, dualism is false.

Since zombies are entirely physical (i.e., everything about their minds logically supervenes on their token physical properties), if they are conscious, consciousness must be physical. Of course, this means that we've mistakenly conceived of zombie twins: they aren't lacking consciousness at all. One might object here that the very definition of "zombie twins" means they can't be conscious. But as we have already seen, the conscious essentialist path through the Contradictory Argument requires changing the definition of "zombie twins." Something has to give.

What gives is the notion that zombies are not conscious. What remains is the idea that zombies are physical twins. This shouldn't be too surprising, since conscious essentialism is assumed in this part of the Contradictory Argument.

Another way to view the situation with premise 7 is to note that Premise 7 has a dual:

7D: "If zombie twins are conscious, zombies aren't entirely physical."

The difference between premise 7 and 7D is this. Ultimately, each path through the Contradictory Argument is designed to *teach* us something about *consciousness*, not zombie twins. 7D teaches us something about zombie twins. But since we are assuming conscious essentialism for this path, we don't need to be taught anything about zombie twins, we already know that there can't really be zombies twins. Hence, if they are conscious, dualism must be false -- they aren't zombies.

This completes our justifications of the premises. The justifications, no doubt, raise further issues, but they are sufficiently strong to make the premises plausible, at least for the nonce. But now we are saddled with the conclusion that dualism is both true and false. Even if one accepts that there are true contradictions (Priest, 2003, 2006), trying to avoid a contradiction here is eminently reasonable. Unfortunately, reasonable though it is, avoiding contradiction is not possible here. This is the matter to which we now turn.

4. Conscious Essentialism and the Impossibility of Zombie Twins . . . and the Return of the Zombies

In this section, we argue for conscious essentialism and embrace its conclusion that zombie twins are impossible. Then we show that zombie twins still have to be possible, if dualism is true, which we also argue is a serious possibility.

4.1. Impossibility of Zombie Twins[8]

Frankly, to us, premise 3 seems obviously true. But Chalmers flatly denies it. He has to deny it because he uses zombie twins to argue for dualism (1996), and by definition, they have to behave exactly like we do – this is captured in the definition that is premise 2. For Chalmers, then, *X* can conceive of its zombie twin and yet not be conscious. So, our zombie twin thinks that it is *not* the zombie twin, but instead, considers *its* zombie twin, for this is precisely what we do. How could our zombie twin think that it's not a zombie? Apparently, it thinks it's conscious, even though it's not.

In chapter 5 of his 1996, Chalmers goes to great lengths to point out and then wrestle with the problem that, on his theory, zombie twins will judge that they are conscious (and judge that they are seeing red, hearing music, etc. etc.). Chalmers's zombie twin will spend large quantities of time working feverishly on a book on consciousness, which requires contemplating his (the twin's) zombie twin (the twin's twin) (1996, p. 180). This seems to be an unhappy conclusion. But it is a conclusion: We judge that we are conscious, so our zombie twins have to, too. Call these *phenomenal judgments*. Our phenomenal judgments flow from our beliefs about our phenomenal experiences: "that is red," "that is the sound of a trumpet," etc. Call these *phenomenal beliefs*. Phenomenal judgments are what you get when you take a phenomenal belief and remove any phenomenal quality (the qualia) (see, Chalmers, 1996, 174). Zombie twins can make phenomenal judgments (according to Chalmers), but cannot have phenomenal beliefs. But now we have an obvious problem: how can our zombie twins make phenomenal judgments about their "experiences" (scare quotes required) when they don't have any – when they aren't even conscious?

Chalmers calls this problem the *paradox of phenomenal judgment* (1996, ch. 5, see, esp., p. 177). Little noted is that this paradox is

---

8 Much of the basic material for this section is taken from Rose (2009).

ambiguous between a positive version involving us and a photo-negative version involving our zombie twins. The positive version of the paradox is this:

> Given that dualism is true, how can physical beings such as we humans have phenomenal beliefs and make phenomenal judgments when the information we need for such mental states is not physical at all?

The negative version of the paradox is:

> Given that dualism is true because zombie twins are possible, how can they ever make a judgment involving phenomenal experience when, in their world, there are no phenomenal experiences (or phenomenal information) at all?

The positive version asks how can non-physical, non-material experience affect our judgments, which are physical (being the result of brain processes) – How does the physical/nonphysical handshake occur? The negative version asks how can purely physical beings make phenomenal judgments when, in their world, the information needed is simply not present – How can there be a one-handed handshake? (a Zen "answer" is inappropriate here, of course).

Chalmers tackles the positive version (1996, ch. 5; and see esp., 2003).[9]

He also attempts to provide a solution to the negative version (1996, ch. 5, section 3). He argues that phenomenal judgments flow solely from cognition (which is completely physical), and that real consciousness is not needed at all. He says: ". . . consciousness is surprising, but claims about consciousness are not" (p. 186). His argument assumes the existence of a

9 In tackling the positive version, he produces an interesting proposal for how the physical/nonphysical handshake occurs. He also surveys in detail the consequences of this theory for minds, their mental states, concepts, representations, and epistemology. For this, see Chalmers, 2003.

computational autonomous agent. However, the argument shifts disconcertingly. When we ask the computational agent how it knows it sees a red tricycle, the agent says "I just see it" (p. 185). So, it seems as if either Chalmers just asserts that the computational agent would make phenomenal judgments without consciousness, or Chalmers implicitly assumes that the agent is conscious at the beginning of the argument, and then jettisons consciousness for the conclusion of the argument. Either way, the argument fails. If consciousness is surprising, then so must be claims about consciousness.

The real problem is that embracing conscious inessentialism is not a solution, it's a consequence of what should be a solution. One cannot just say that our zombie twins (or other unconscious agents) make phenomenal judgments; one has to provide an account of *how* they will make their judgments without consciousness. This is because the strong belief to the contrary must be overcome. It is very hard to believe that phenomenal judgments don't require phenomenal experiences, i.e., conscious *essentialism* is very easy to believe, indeed, it is the natural, default belief. But worse, phenomenal judgments connect smoothly with the rest of our mental lives – to phenomenal beliefs, specifically. Much of our mental lives is profoundly informed by our conscious experiences. We talk about consciousness because we are conscious – what could be more obvious? It is completely baffling how zombie twins could talk about consciousness. So, how could zombie twins have anything like the mental lives we have?

To get a sense of how strange this is, note that Chalmers's zombie twin produced an argument for dualism and published it. In fact, getting all agitated over the nature of consciousness doesn't even require consciousness to exist! Suppose that consciousness never existed in the first place; the universe only had zombies in it (what would have been our zombie twins had we existed). Then those zombies would still be able to prove that dualism is true. Dualism might well be true, but is bizarre to think that it could be proved true in a universe devoid of consciousness. One cannot just label all these cases of conscious inessentialism and move on; this problem cries out for a substantive solution for how our zombie twins could think, say, and do exactly what we do. But there is no solution. There's no way to

explain how zombies can talk about consciousness, or the color red, or the sound of a trumpet, etc. if they aren't conscious.

At this point, one can conclude that zombie twins won't have any mental states at all similar to ours, since their states are not remotely connected to conscious experiences. Plausibly, they neither judge nor believe that they are conscious. Fish don't dream of climbing Mount Everest. It is not that zombie twins judge *incorrectly* that they are conscious, rather, zombie twins don't think about consciousness in any way at all. But then zombie twins aren't much like us. This is just another way of saying that zombie twins are *impossible*: they aren't our *twins*. Which in turn is conscious essentialism. We can conceive of zombie twins, but only in a rough, crude, or superficial way, similar to the way we conceive of round squares.[10]

We conclude that the notion of zombie twins is unworkable, and probably incoherent. Any such "being" would either be not a twin or not a zombie. So there are no zombie twins. The conscious inessentialism path through the Contradictory Argument is not a viable path at all. So the Contradictory Argument is defused.

4.2. The Return of the Zombie Twins

Yet, zombie twins are possible. So, the Contradictory Argument is reinstated. Here's how this comes about.

Even setting zombie twins aside, we have other, very good reasons to believe that dualism is true. Inverted spectra are one such reason. Though zombie twins are impossible to conceive in any useful detail, it is far easier to imagine inverted twins. One's inverted twin perceives an inverted color

---

10 Using Kripkean modal definitions and arguments, Dietrich and Gillies (2001) argue that zombie twins cannot be conceived in the way required for Chalmers' dualism argument. The only way to pick out a twin of some conscious being in another world, without begging the dualism/materialism question in favor of dualism, is to use essences (haecceities), and consciousness is the only essence in the vicinity. So there is no possible world where, e.g., David Chalmers is not consciousness – such a being wouldn't be David Chalmers.

spectrum relative to you. One's inverted twin sees yellow where you see blue, and vice versa. In this case, most of the conceptual and logical problems that plagued zombie twins vanish: inverted twins *are* conscious, they see color, it is just that their experiences are systematically different. All the physical facts about you are true of your inverted twin, but the phenomenal facts are different. This difference is sufficient to guarantee that phenomenal facts don't superdupervene on the physical. Hence, dualism is true. But if dualism is true, then zombie twins are logically possible -- i.e., there exists a possible world with the same physical facts as the actual world, but no phenomenal facts at all, for to insist otherwise seems to tie the phenomenal to the physical in a way that requires superdupervenience, which would mean that dualism is false. So if dualism is true, zombie twins are possible, and dualism seems true. Hence, the conscious inessentialism path through the Contradictory Argument returns, alive and well. [11]

This result is exceedingly disconcerting. Conscious essentialism seems not just true, but obviously true; zombie twins are right out; they are impossible. Yet, dualism appears true for other, non-zombie reasons. And if dualism is true, then since this entails that consciousness doesn't superdupervene on the physical, zombie twins are apparently possible after all. It seems as if the only conclusion has to be both that zombie twins are not conceivable, but possible (conscious inessentialism), and also not possible (conscious essentialism). We locate the source of this problem not in zombies, nor in inverted twins (or conceptions of them), but in consciousness itself. When thinking about oneself, one's experiences, and one's *knowledge* of such experiences, consciousness is revealed as essential. But when thinking about others and their knowledge and experiences, consciousness emerges as inessential (or at least conceivably inessential) because others' knowledge and experiences are accessible only via overt behavior, and this behavior apparently can remain invariant under wildly

---

11 With some extra work, this same result could be established with Jackson's Mary argument, which is an epistemological argument showing that knowing all the relevant physical facts does not entail knowing any phenomenal facts at all. See Jackson, 1982.

differing conscious experiences. It is, it seems, a small leap from wildly differing conscious experiences (e.g., inverted twins) to no experiences at all (zombie twins). The big, and crucial, leap is from focusing on one's self (an inner focus) to focusing on others (an outer focus). This shift between inner and outer infects another long-standing and important debate in metaphysics: the realism/anti-realism debate. We will argue that this debate has the same structure as the zombie twin debate, with identical consequences: realism and anti-realism are both true, just from different points of view, both of which enjoy equal status as the correct point of view. And again, consciousness is the culprit.

5. Realism versus Anti-Realism: It's All Points of View in the Void

We define *realism* as the thesis that there is a mind-independent world. *Anti-realism* is the denial of this: there is no mind-independent world. (Here, we ignore further restrictions that can be placed on these definitions.) Realism and anti-realism are equally true. By this, we mean that realism is true from one point of view and anti-realism from another, and both points of view have equal and legitimate claims to being the preferred point of view. This situation is due to consciousness's property of engendering points of view. We argue for all this, in this section.

Rather than beginning with realism and anti-realism, we begin with the two points of view we are interested in.[12] We dub these: the *view of no one*, and *solipsism* (the view of exactly one).[13]

Solipsism is the view that everything is mind-dependent. All that really exists is the mind of the solipsist, *S*. Everything else exists only as the experiences of *S*, including *S*'s body. All people, things, processes have their being only as conscious experiences of *S*. Solipsism is an ontological

12 The way of couching the material of this section is derived from Hannah Rice's paper "You simply cannot think solipsism is true" (2009).

13 See Dietrich, 2008, for details on the view of no one. Also, there, the view of no one is used to construct a true contradiction – a dialetheia.

thesis, based on a epistemological foundation: we only have experiences, and only their phenomenological character is epistemically certain, and only what is certain is knowable. Solipsism is profoundly anti-realistic.

The view from solipsism is easy to adopt. Almost everyone has wondered if it is true. All the available evidence is compatible with its truth. (Viewing the movie *The Matrix* is a good introduction to a pre-solipsistic, reality-equals-just-what-we-experience point of view. From there, solipsism is easily attained.)

From the view of no one, there is no mind-independent world because there are no minds – no one has a mind at all. Everything is entirely *world-dependent*, as it were. What really exists are physical beings, things, and processes. Minds are only an illusion (a delusion, actually). *A fortiori*, consciousness is a delusion. No one ever has qualitative experiences; no one ever has had qualitative experiences.

The view of no one is somewhat hard to adopt. Interestingly, it is the basis of a branch of Buddhism. The best introduction to the view of no one is Douglas Harding's *On having no head* (1972):

> The best day of my life -- my rebirthday, so to speak -- was when I found I had no head. This is not a literary gambit, a witticism designed to arouse interest at any cost. I mean it in all seriousness: I have no head.
>
> It was eighteen years ago, when I was thirty-three, that I made the discovery. Though it certainly came out of the blue, it did so in response to an urgent enquiry; I had for several months been absorbed in the question: what am I? The fact that I happened to be walking in the Himalayas at the time probably had little to do with it; though in that country unusual states of mind are said to come more easily. However that may be, a very still clear day, and a view from the ridge where I stood, over misty blue valleys to the highest mountain range in the world, with Kangchenjunga and Everest unprominent among its snow-peaks, made a setting worthy of the grandest vision.
>
> What actually happened was something absurdly simple and unspectacular: I stopped thinking. [. . .] There existed only the Now . . . To look was enough. And what I found was khaki trouserlegs terminating downwards in a pair of brown shoes, khaki sleeves terminating sideways in

> a pair of pink hands, and a khaki shirtfront terminating upwards in -- absolutely nothing whatever! Certainly not in a head.
>
> It took me no time at all to notice that this nothing, this hole where a head should have been, was no ordinary vacancy, no mere nothing. On the contrary, it was very much occupied. It was a vast emptiness vastly filled, a nothing that found room for everything -- room for grass, trees, shadowy distant hills, and far above them snow-peaks like a row of angular clouds riding the blue sky. I had lost a head and gained a world.

Clearly, from the view of no one (the view of having no head), there is only the world. There is no individual experience of it at all. As we mentioned, the view of no one is somewhat difficult to adopt, but it can be adopted for very short intervals of time rather easily. Like the solipsism view, the view of no one is an ontological thesis, based on epistemology: what we really know is the world; it is the world that ultimately exists, not spectators of that world. The view from no one is profoundly realistic: there's definitely a mind-independent world because there is a world, but no minds.

It is possible to sit in one's study and move between solipsism and the view of no one. After a short amount of time practicing, sliding between these two points of view becomes as easy as walking back and forth in a room (if only for short intervals of time). But moving between these two points of view is moving between anti-realism and realism. From the solipsistic point of view, there is no mind-independent world; from the view of no one (no-head), there is *only* a mind-independent world; what doesn't exist are minds. Back and forth we can go. Realism is just a point of view, and so is anti-realism. And both points of view have equal claims on our assent; neither can declare victory. All the evidence available – sensory information, introspected information, and information derived via logical reasoning – is compatible with either point of view. Arguments for either point of view also are equally persuasive.

Solipsism and the view of no one are not the only points of view operative here. We admit that the default point of view for most readers is another kind of realism, roughly in the middle of the between the other two:

there is a mind-independent world perceived by beings with minds. This is a very natural and common point of view, but it, too, is just a point of view along the way between solipsism and the view of no one, and no argument for it trumps either solipsism or the view of no one. It's the default for pragmatic reasons only. We call this point of view *quotidian realism*. There is, also, a fourth point of view from which the first three are viewed; the reader is occupying this viewpoint now. This is a *meta* point of view: from it, points of view are viewed. Primarily, it this meta point of view that makes realism and anti-realism points of view and not rivals for truth, for from the meta point of view it is apodictic that realism and anti-realism are just two different ways of being conscious. More importantly, from this meta view, the contradiction resulting from solipsism and the view of no one is perceived; that is, the contradiction between anti-realism and realism is perceived. Indeed, all three -- anti-realism, realism, and quotidian realism -- are mutually contradictory, and all these contradictions can be seen from the meta point of view.

The fourth, meta point of view shows that points of view are not *interpretations*, if that term is understood to mean a gloss on some point-of-viewless, objective reality, or on some raw, point-of-viewless experience. Furthermore, in the present context, assuming a strong version of the claim that points of view are interpretations would introduce a contradiction into our analysis (a bad contradiction of the only-false variety), since it would be assuming realism or anti-realism; or such assuming would beg the question against us in a challenge to our analysis. Happily, introducing points of view, on the other hand, does not beg any questions in our favor, since they are compatible with either realism as well as with anti-realism, as we will now see.

Points of view are contexts occupied by conscious minds. Two crucial properties of points of view and conscious minds are:

> Necessarily, all conscious minds occupy some point of view or other. (Independently of a point of view, there is no information to be conscious of. This can be construed as being either a property of mind-

independent information (the realistic construal) or as a property of conscious minds (the anti-realistic construal).)

Necessarily, a point of view exists only if it is occupiable by a conscious mind. (These and other aspects of points of view are discussed at length in Julian and Dietrich, 2008.)

Being from a point of view is an essential property of consciousness, just as being even is a essential property of 6. Philosophers and others productively talk about consciousness without mentioning points of view, just as one can discuss 6 without mentioning its evenness (one can, for example, point out that it is one bigger than 5 and one smaller than 7, or that it is the smallest perfect number, a number which is the sum of its proper factors). But philosophers get into no end of trouble when they vie with each over what are, at root, just ways of being conscious. The only truth in the vicinity is that all the relevant ontological positions are true, just from a point of view. So, realism and anti-realism are both true, just from different points of view.

Realism and anti-realism, as we have analyzed them, are contradictory. This is because the two relevant points of view, solipsism and the view of no one, are contradictory. We can get a better handle on this matter if we consider an analogy.[14] Consider Figure 1, a Necker Cube.

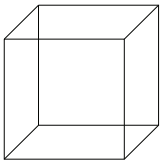


Figure 1. A Necker Cube.

Does it face down and to the left or up and to the right or is it just twelve lines on a flat page?

14 We adapted this analogy from a good objection that Graham Priest made to an earlier version of this paper.

There are three points of view (at least) on the Necker Cube; the first two are the most common. One can see it as a three dimensional rectangular box facing down and to the left or as a box facing up and to the right, or one can see it as a planar figure comprising twelve lines at various angles with one another. (One could also see it as 16 lines a page, or as two triangles and five quadrilaterals all sharing some sides, or as a combination of these two. We'll stick with the first three.) Call the first point of view (down and to the left) DL; call the second point of view (up and to the right) UR; call the third point of view (twelve lines on a flat page) 12L. From these points of view, the specific versions of the Necker Cube are perceived.

12L functions in this analogy just like quotidian realism. We call it *quotidian Necker*. Quotidian Necker, like quotidian realism, is considered by most to be the fundamental truth of the situation. Quotidian Necker is: "There are some lines on a page and the human visual system interprets those lines as a three dimensional box. But given the way the lines are drawn (i.e., there's no indication of occluding), the lines on the page are *ambiguous*: they are interpreted by the human visual system as being a box pointing down and to the left or up and to the right, and the interpretation vacillates between the two." But again, Quotidian Necker is just a point of view; it is not the fundamental truth -- there isn't a fundamental truth, here. To see this, note that there is a fourth point of view from which DL, UR, and 12L are viewed, and from which one can see that these three are points of view and are also mutually contradictory. We call this fourth view, 4V. From 4V, a contradiction is perceived, a three-way contradiction, in fact. And from 4V, the truths perceived from DL, UR, and 12L are all equally plausible; no point of view trumps the others. Hence, from 4V, the contradiction between DL, UR, and 12L is genuine.

What we'd like to do at this juncture is conclude that consciousness itself is contradictory, and that it is this that explains the contradictions between realism and anti-realism and between conscious inessentialism and essentialism. Unfortunately, this conclusion is unwarranted at this time.[15] We know only that consciousness admits of contradictory viewpoints and

---

15 Graham Priest pointed this out.

that these viewpoints are necessarily tied to consciousness. From this, we cannot conclude that consciousness itself is inherently contradictory.

But we can get close. Consider the Necker Cube again. By themselves, neither DL, UR, nor 12L are contradictory. They are only contradictory in pairs. One naturally seeks an explanation of this situation, and when one does that, there is a strong tendency to deny that all three are just points of view, equal in status, and instead to claim that the fundamental object here is 12L, and it is inherently ambiguous, but not contradictory. Just so, one might dig in one's heels and claim that the same is true of realism, anti-realism, and quotidian realism. Quotidian realism is the fundamental truth, it is just that reality with conscious minds in it is ambiguous between realism and anti-realism. But which is it? Is it reality that is ambiguous or do conscious minds produce the ambiguity? The background assumption here doing all the work is that some *one thing* needs to be responsible for the contradictory nature of realism and anti-realism. If it is reality that is ambiguous, that is hardly in keeping with what we might call the "spirit of realism," for in this case, there really isn't a mind-independent world -- there is, rather, some "mind-independent" ambiguous stuff (perhaps it is noumenal). If it is the conscious mind that is ambiguous (or if it is consciousness itself), that is hardly in keeping with what we might call the "spirit of anti-realism," for in this case, there really isn't a mind-dependent world -- there is, rather, some ambiguous "thinking" stuff which sometimes reveals solipsism to be true and sometimes produces a "mind-independent" world (which could be contradictory, depending on what "thinking" turns out to be in this context). Neither option is acceptable -- neither reality nor the conscious mind can be ambiguous while preserving the basic character of either realism or anti-realism. Given this, concluding that consciousness is inherently contradictory gains some credibility.

But perhaps we should throw out the background assumption that one thing needs to be responsible for the contradictory nature of realism and anti-realism. Perhaps what is ambiguous is the world with conscious minds in it. We could even legislate this to be one thing by adding hyphens: the ambiguous thing is the world-with-conscious-minds-in-it. But does this mean that before there were minds in the world, the world wasn't

ambiguous? Unfortunately, this question is illegitimate here since in presupposes realism.

But even granting that the world-with-conscious-minds-in-it is ambiguous doesn't help much. The *ambiguity thesis* is that what explains the unresolvable contradiction between realism and anti-realism (between what's perceivable from the view of no one and from solipsism) is that something is inherently ambiguous. The essential problem with this thesis is that it appears to require an ultimate reality: the thing that is ambiguous. This is question-begging in the present context, for though it violates the "spirit of realism," the ambiguity thesis is nevertheless enough of a realism to beg the question, here. Furthermore, there is a good argument against the ambiguity thesis. This is the argument we presented when we introduced the fourth points of view: meta and 4V. From these points of view, that everything is a point of view is readily perceived, along with their ineluctable contradictions. So, it appears, the ambiguity thesis is out. But ambiguity and contradictory consciousness seem to be the only candidates on offer. If so, then it is plausible that consciousness is inherently contradictory.[16]

Here is what we've got. Either consciousness is inherently contradictory or the world with minds in it is inherently ambiguous. A good case can be made that it is consciousness that is inherently contradictory. In any case, both realism and anti-realism are here to stay. And so are conscious essentialism and inessentialism. And either way, consciousness is heavily implicated. Consciousness, whatever it is, is the sort of thing that allows . . . encourages . . . *causes* (?) . . . contradictory points of view. And perhaps this explains not just why ontology and metaphysics are so perplexing, but why all of philosophy is.

---

16 In dialetheic contexts, and in paraconsistent logic in general, the argument form *disjunctive syllogism* is not in general valid. We aren't using disjunctive syllogism here, for we aren't making a deductive argument, but, rather, an inductive one.

Appendix

A proposition, *P*, is *conceivable* if it can be brought before the mind. This is often (but not always) done by bringing before the mind some situation in which *P* is true. Another way is looking for but not finding any contradiction in, or entailed by, *P*. *P* is *ideally* conceivable when the conceiving of *P can't* be undone by better reasoning. For example, suppose that, someone, say Girolamo Saccheri conceives that Euclid's fifth postulate (the parallel postulate) is derivable from the other four. A better reasoner comes along, say, Riemann, and demonstrates that the parallel postulate is independent of the other four. This shows that though one can conceive of proving the parallel postulate from the other four, one cannot ideally conceive this. *P* is *positively* conceivable when one can bring before one's mind a situation in which *P* is true. This definition rules out one type of basic conceivability: negative conceivability. *P* is *negatively* conceivable when it is not ruled out, *a priori*. Positive conceivability, by contrast, actively rules something in. Finally, *P* is *primarily* conceivable when it is conceivable that *P* might *actually be* the case. This contrasts with *secondary* conceivability, which is conceiving of *P* subjunctively, i.e., as what *might have been* the case. (All of these definitions come from Chalmers, 2002. See also, Chalmers, 1996, ch. 2.)

So now we have defined ideal, primary, positive conceivability. In sum, it is conceiving a situation in which *P* is actually the case, and where such conceiving cannot be undone by better, more thorough conceiving.

As with conceiving, there are varieties of possibility (Chalmers, 2006). The only one we will need is *primary possibility*. First, the kinds of possible worlds we will use (following Chalmers) are *centered* possible worlds (1996, 2002). Centered worlds have a central point of view or focus within them. The point of view is that of a specified or privileged agent in that world. Centered worlds are required to handle issues involving indexicals, which clearly arise when the topic is consciousness, for consciousness is indexical: each of us knows only his or her consciousness. Next, and again following Chalmers, the *primary intension* of a proposition *P* is a function that takes *P* and a world *W* as input and returns the truth value of *P* at *W*, where *W* is considered as actual, rather than counterfactual

(2002). Another way to run the definition is to use the notion of *a priori* entailment. This gives: the primary intension of *P* is true at W if the material conditional "if *W* is actual, then *P*" is *a priori* true (2002). Consider the well-known proposition "Water is XYZ." (XYZ is an alternate chemical nature of water -- that is, the clear, drinkable, life-sustaining stuff in rain, streams, oceans, etc. -- in the XYZ possible world; XYZ is not H_2O.) If the XYZ world is considered as actual, then the primary intension of this proposition is *true*. "Water is H_2O" is also true in any H_2O-world, using the primary intension. Kripke's famous insight that it is a necessary *a posteriori* truth that water is H_2O obtains only for the *secondary intension* of "Water is H_2O." The secondary intension of *P* takes *P* and *W*, *considered as counterfactual*, and returns the truth value of *P* at *W*. So, given that water is H_2O, i.e., that science has revealed this fact, then it's false that water is XYZ in the XYZ world (or, if one likes, in any XYZ world), since H_2O is not XYZ. Yes, there's some sort of clear, drinkable stuff in the streams of the XYZ world, but it is not water (2002).[17] As mentioned, primary intensions are known *a priori*; secondary intensions are *a posteriori*. Now, to complete the definition of primary possibility: *P* is primarily possible when its primary intension is true in some possible world considered as actual.

The tight connection between ideal primary positive conceivability and primary possibility should start to be apparent. The secondary intension of "Water is XYZ" is true in no possible world. Considered counterfactually -- that is, where water is in fact H_2O -- then whatever XYZ is, it's not water. But we do conceive of water being XYZ (we've have done so several times,

---

17 This analysis relies on the more basic notions of the primary and secondary intensions of a concept. The primary intension of a concept does not depend on the world science reveals to us (Chalmers, 1996, 2002). Rather, it depends on how reference is fixed in the actual world from the point of view of the subject. So, the primary intension of the concept "water" is (roughly) the clear, drinkable stuff which is required for life and is found in our lakes, streams and oceans (Chalmers, 1996, ch. 2). Given that water is revealed to be H_2O, the *secondary intension* of the concept "water" is H_2O. Hence, the secondary intension of "water" picks out the water (the H_2O) in all counterfactual worlds. This all forms part of Chalmers's *two-dimensional model* of modal semantics (see, esp., 2006, and also 1996, 2002).

here). This conceiving is of a different sort; it relies on conceiving what might actually be the case. It therefore relies on primary intensions. The primary intension of "Water is XYZ" is true in those XYZ centered worlds considered as actual. Translated, this proposition says "The clear, drinkable, life sustaining stuff found in oceans and streams is XYZ." Clearly, this is conceivable (primarily), and so conceived, there is a possible world where it is true, namely, the XYZ world (2002).

Now, we have:

> Ideal primary positive conceivability entails primary possibility (2002).

Or, to paraphrase Chalmers: If a proposition *P*, is ideally, primarily, positively conceivable, then there is a metaphysically possible centered world, considered as actual, where *P*'s primary intension is true (2002). This seems quite plausible given the discussion above: both the relevant conceivability and possibility are based on the fundamental notion of a primary intension (of conceiving for the antecedent, of possibility (or the truth in a possible world of a proposition) for the consequent). This ties the two together so closely that the truth of the former entails the truth of the latter.

Conceivability might imply possibility using other forms of conceivability and possibility (Chalmers, 2002). But be that as it may, this is the only case we need for premise 1. For, if some conscious agent, *A*, ideally, primarily, and positively conceives of its zombie twin, then it is conceiving, in a way that cannot be undone, of a situation where the physical facts of the actual world obtain without consciousness thereby obtaining.[18] In short, *A* conceives of the zombie world as actual. But in that

---

18 Technically, this is saying that consciousness doesn't logically supervene (using Chalmers's notion) on the physical facts of our universe. Which in turn means that consciousness is a further, extra fact about our world. Which in turn means that materialism is false and some sort of dualism (at least) is true. We turn to this shortly. See, Chalmers, 1996.

possible world, *A*'s crucial proposition, "I have a zombie twin," is true, *a priori*, as required.

Here's another angle on this using the first-person indexical (the reader is requested to put him/herself in for all the first-person terms). Given that I ideally, primarily, and positively conceive of my zombie twin, the question becomes "Is 'I have a zombie twin' primarily possible?" This latter, in turn, is the question "Is the primary intension of 'I have a zombie twin' true when evaluated at the zombie world, when that world is considered as actual?" The answer is clearly, Yes. (Remember, we are assuming, because we have to here, that conscious inessentialism is true. The conscious essentialist will deny that anyone can positively conceive of his or her zombie twin. Or the essentialist will deny that ideally conceiving of one's zombie twin is impossible. Or both.)

Primary intensions dominate the situation, here, because secondary intensions are useless: we don't know what consciousness could be counterfactually, since we lack an analysis of it (scientific or otherwise). So, conceiving that "I have a zombie twin" might actually be the case guarantees that there is a possible world where "I have a zombie twin" is true. One might put the matter this way: The positive situation conceived when conceiving of one's zombie twin *just is* the relevant zombie world; the very zombie world that is conceived in the antecedent is accessed in the consequent. So, of course, (ideal . . .) conceivability implies (primary) possibility.

Now to establishing dualism. We begin with the definition of logical superdupervenience:

> A facts logically superdupervene on B facts iff any two logically possible worlds identical in their B facts are *eo ipso* identical in their A facts, and the A facts are robustly explainable in terms of the B facts because of the "*eo ipso*" condition.

Everything in the world logically superdupervenes on the level below it. Fix the low-level physical facts of our world, the behaviors and trajectories of every particle -- every quark, electron, proton and neutron --

and you automatically fix all the other facts in our world -- the chemical facts, the biological facts, the psychological facts, and the social and cultural facts. In other words, it is logically impossible to for there to be a world just like ours at the lowest level, that has exactly the same detailed, low-level physical facts as our actual world has, but which differs from our world in its high-level facts. Hence, it is impossible to ideally, primarily, positively conceive of such a world.

Here's an example using a glass of water. Conceive of a glass filled with hot water. The atoms in the glass are caroming all over the place in a very agitated way. Now, try to conceive of another glass of water where the atoms are behaving in *exactly* the same way as in the first glass, but where the water in the second glass is cold. You can't do it. Or, if you think you can, you are mistaken (c.f., Chalmers, 1996, p. 109). For, all we *mean* by "hot" is that the atoms are caroming all over the place in a very agitated way. Fix the behavior of the water atoms in the glass and you automatically fix the water's temperature. This example exhibits just what is going on at the level of our entire universe. It is simply inconceivable that the low-level facts about our world could be what they are and yet there be no stardust, no suns, no galaxies, no planets, no continents, no minerals, no life, no US Constitution, no penguins in Antarctica, and no MTV (the Music Television Channel). In short, and though it may sound strange, MTV logically superdupervenes on the low-level physical facts of our world. There is no possible world with the same low-level facts as ours that isn't blessed with MTV. This superdupervenience hierarchy subsumes everything[19]; everything in our world superdupervenes *logically* on the level below it and ultimately on the lowest level -- everything, that is, but consciousness, which we know doesn't *logically* superdupervene since there is a possible world, the zombie world, where the physical facts of our world obtain, yet there is no consciousness. So consciousness is revealed as an *extra* fact in

19 There are some technical tweaks that have to be made to make this statement true. We will skip those. Chalmers handles them fully in chapter 2 of his 1996.

our world, a fact that is not guaranteed by the physical facts. Hence, dualism is true.

This completes our defense of premise 1.

References

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory.* Oxford: Oxford University Press.

Chalmers, David (2002). Does Conceivability Entail Possibility? In Tamar S. Gendler & John Hawthorne (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press.

Chalmers, David (2003). The Content and Epistemology of Phenomenal Belief, in *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Alexandr Jokic. Oxford: Oxford University Press.

Chalmers, David (2006). Two-dimensional Semantics, In E. Lepore & B. Smith, eds. *Oxford Handbook of the Philosophy of Language*. Oxford: Oxford University Press.

Chalmers, David J. & Jackson, Frank (2001). Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110 (3):315-61.

Dennett, D. (1988). Quining Qualia, in A. Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*, Oxford University Press.

Dennett, D. (1995). The unimagined preposterousness of zombies. *J of Consciousness Studies,* 2 (4), pp.322-326.

Dietrich, E. (2008). The Bishop and Priest: Toward a point-of-view based epistemology of true contradictions. *Logos Architekton*, v. 2, n. 2 pp. 35-58.

Dietrich, E. and A. Gillies (2001). Consciousness and the limits of our imaginations. *Synthese* v. 126, n. 3, pp. 361-381.

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Flanagan, O. and T. Polger (1995). "Zombies and the Function of Consciousness." *Journal of Consciousness Studies* (1995), 2(4):313-321.

Harding, Douglas. E. (1972). *On Having No Head*. Perennial Library, Harper and Row. First published in 1961.

Horgan, T (1993). From Supervenience to Superdupervenience: Meeting the Demands of a Material World, *Mind*, vol. 102, No. 408. pp. 555-586.

Julian, J. and E. Dietrich (2008). Points of view from a novel perspective, in S. Scalet, (ed.) *Social Philosophy and Our Changing Points of View*. Binghamton, NY: Global Academic Publishing.

Moody, T. C., (1994). Conversations with zombies, *J. of Consciousness Studies*, 1 (2), pp. 196-200.

Murphy, G. (2002). *The Big Book of Concepts.* Cambridge, MA: MIT Press.

Pepperberg, Irene Maxine (2000). *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots.* Cambridge, MA: Harvard Universitiy Press.

Priest, G. (2003). *Beyond the Limits of Thought*. 2nd ed. Oxford: Oxford University Press.

Priest, G. (2006). *In Contradiction: A Study of the Transconsistent*. 2nd ed. Oxford: Oxford University Press, 2006.

Rice, H. (2009). *You simply cannot think solipsism is true.* Unpublished Manuscript. Binghamton University.

Rose, J. (2009). *Is Chalmers's Zombie Argument Self-refuting? And how.* Honors Thesis, Dept. of Philosophy, Binghamton University, Binghamton, New York.

Valdman, M. (1997). Will zombies talk about consciousness? – The paradox of phenomenal judgment: its implications for a naturalistic dualism and other theories of mind. *J. of Experimental and Theoretical AI*, 9, pp. 471 – 490.

Realism, Relativity and Representation

Friedel WEINERT [*]
University of Bradford (UK)

Abstract:

The paper argues that Einstein's distinction between '*constructive* and *principle* theories' involves representational claims about physical reality and therefore has implications for the question of realism. Einstein was mostly interested in the latter kind of theory because it imposes fundamental *constraints* on both the phenomena and their scientific representation. The Special Theory of Relativity (STR) represents physical reality in such a way that only the invariant is to be regarded as physically real. This invariance view arises from the imposition of constraints on the reference frames in the STR. A consideration of constraints shows that structures are of central concern in the relativity theory. The concern for structure puts Einstein's views in the vicinity of structural realism.

Keywords: Constraints, Invariance, Perspectivalism, Realism, Reference Frames, Relativity, Representation, Structure, Symmetry

Introduction

The line of argument pursued in this paper is to proceed from Einstein's fundamental problem situation to a consideration of scientific representation with respect to the Special theory of relativity (STR). Einstein's fundamental problem situation, which is Kantian in spirit, is how the conceptual freedom of the scientist is compatible with the need for an objective representation of an independently given material world. To solve this philosophical issue Einstein employs a number of constraints, which are central to the STR. The issue of scientific representation leads to a

[*] E-mail: F.Weinert@bradford.ac.uk.

consideration of the notion of reality and to the realistic commitments implied in the STR. From this point of view, the paper concludes that Einstein was committed to a kind of 'structural' realism.

Concepts, Facts and Constraints

1. Einstein's fundamental philosophical position arises from the age-old puzzle of how concepts are related to facts. More generally, how do scientific theories represent empirical reality? Einstein warned against the tendency to regard concepts as thought necessities. Once certain concepts have been formed, often on the basis of experience, there is a danger that they will quickly take on an independent existence. People are tempted to invest them with some kind of Kantian necessity. Concepts, however, just like theories, are always subject to revisions. Einstein complained that

> Philosophers had a harmful effect upon the progress of scientific thinking in removing certain fundamental concepts from the domain of empiricism, where they are under our control, to the intangible heights of the a priori. (Einstein 1922, 2)

What Einstein had in mind were the classical notions of space and time. Newton had regarded it as necessary to introduce the notions of absolute and universal space and time into his mechanics in order to make sense of his laws of motion. These notions had become part and parcel of classical physics. Kant turned them into thought necessities, although in his *Critique of Pure Reason* he rejected the Newtonian view that space and time had an existence outside of the human mind. The Special theory arrived at a different result. Temporal and spatial measurements became relativitized to particular reference frames. This was a necessary consequence of embracing the principle of relativity and taking the velocity of light as a fundamental postulate of the theory. Through his own work Einstein had witnessed how such fundamental philosophico-physical notions as space and time required conceptual revision. This made him forever suspicious about the sway that such notions could hold over the minds of physicists and philosophers.

As is well-known Einstein characterizes scientific theories and the fundamental notions of physics - energy, event, mass, space, and time - as free inventions of the human mind. No amount of inductive generalizations can lead from empirical phenomena to the complicated equations of the theory of relativity. Science, however, assumes the existence of an external world. Furthermore, scientific theories are meant to entail objective statements about the external world. Although the fundamental notions of physics are logically speaking free inventions of the human mind, they must be mapped onto the data given by empirical reality through experiments and observation. (Einstein 1920, 141) Thus Einstein faced the fundamental Kantian position of finding a synthesis between reason and experience. This problem situation poses the question of scientific representation. The notion of *constraint* is of particular importance in an assessment of how the theory of relativity deals with the representational link between concepts and facts, between models and physical systems.

In an article written for the London *Times* Einstein introduces the now famous distinction between *constructive* theories and *principle* theories. (Einstein 1919) Constructive theories employ relatively simple formalisms, which are meant to represent the hypothetical structure of a physical system. The role of a constructive theory is to propose hypothetical (or *as-if*) models, which assign an underlying structure to the observable phenomena. The hypothetical structure is meant to explain the observable phenomena. The kinetic theory of gases models the behaviour of gas molecules *as if* they were billiard balls. Early atom models modelled atoms *as if* they were tiny planetary systems. A constructive theory, in order for its models to represent the observable phenomena, introduces in its formalism a number of idealizations and abstractions. The models represent the phenomena *as if* they only consisted of the components, which the model introduces. Nevertheless, for the representation to succeed the models must retain a degree of approximation to the systems modelled.

Einstein was mostly concerned with theories of principles. Principle theories employ very general features of natural systems, from which mathematical criteria follow, which natural events and their models must obey. The role of a principle theory is to propose well-confirmed

fundamental physical principles: the laws of thermodynamics, the principles of relativity, of covariance and invariance, and the constancy of light. These principles forbid the occurrence of certain physical events, like the propagation of signals beyond *c* or perpetual motion machines. They constitute constraints on the construction of models and theories and the postulation of laws of physics. Constraints can be understood as restrictive conditions, which such symbolic constructs must satisfy in order to qualify as admissible scientific statements about the natural world. Principle theories seek to represent physical systems under the constraint of these principles. If principle theories differ from constructive theories, it is to be expected that they employ more sophisticated models to represent aspects of the external world.

Einstein implicitly talks about various kinds of models, associated with constructive and principle theories respectively. Theories seem to represent via different kinds of models; in this representation, different kinds of constraints seem to be involved. The idea that constructive and principle theories represent different aspects of natural systems raises immediate questions about realism. To which extent can the models of the theory of relativity be regarded as realistic representations of natural systems? Attention should be directed to a number of constraints, which arise from the theory of relativity. The focus on constraints implies a view of scientific representation: that representation is a question of fit. Einstein hints at a notion like 'fit'.

> We have thus assigned to pure reason and experience their places in a theoretical system of physics. The structure of the system is the work of reason; the empirical contents and their mutual relations must find their representation in the conclusions of the theory. In the possibility of such a representation lie the sole value and justification of the whole system, and especially of the concepts and fundamental principles which underlie it. (Einstein 1933, 272)

A scientific theory constructs a coherent and logically rigid account of the available empirical data. Its coherence may always come under threat

with new empirical discoveries. There is nothing final about the representation of a scientific theory of the external world. In his philosophical writings Einstein insists on the logical simplicity of a theory and testability as constraints to be imposed on admissible scientific theories (Einstein 1949a, 22) Logical simplicity is a *methodological* constraint. Compatibility with available and new evidence is an *empirical* constraint.

Although Einstein claims that 'the world of phenomena uniquely determines the theoretical system' (Einstein 1918b, 226), it is clear from a study of the theory of relativity that, apart from 'external confirmation' and 'inner perfection', further constraints come into play. Einstein sees the importance of principle theories in the introduction of fundamental principles – like the relativity principles – which act as constraints or limiting principles. (Einstein 1920, 99; Einstein 1950, 352) For instance, he speaks of the requirement that the laws of physics must be invariant 'with respect to the Lorentz transformations':

> This is a restricting principle for natural laws, comparable to the restricting principle of the non-existence of the *perpetuum mobile* which underlies thermodynamics. (Einstein 1949a, 56)

Einstein holds that the interplay of such constraints – and others like covariance, invariance– creates a *fit* of the theory or model with the evidence extracted from the external world. (Einstein 1949a, 23; Einstein 1918b, 226, Einstein 1944, 289) The representation is described in terms of fit, which is understood in terms of satisfaction of constraints. A theory 'represents' a section of the empirical world, if it satisfies a certain number of constraints.

> In order that thinking might not degenerate into 'metaphysics', or into empty talk it is only necessary that enough propositions of the conceptual system be firmly enough connected with sensory experiences and that the conceptual system, in view of its task of ordering and surveying sense-experience, should show as much unity and parsimony as possible. (Einstein 1944, 289; Einstein 1949b, 669, 680)

The representation is not an image, nor need it be perfect or absolute. Fit is a matter of degrees. It changes with the changing nature of constraints. In his discussion of principle theories Einstein explicitly states that such theories employ principles 'that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy.' (Einstein 1919, 228) The more constraints are imposed on scientific constructs, the greater the chance that representation will succeed.

> The physical world is represented as a four-dimensional continuum. If I assume a Riemannian metric in it and ask what are the simplest laws which such a metric can satisfy, I arrive at the relativistic theory of gravitation in empty space. If in that space I assume a vector-field or an anti-symmetrical tensor-field which can be derived from it, and ask what are the simplest laws which such a field can satisfy, I arrive at Maxwell's equations for empty space. (Einstein 1933, 274)

2. Let the empirical facts, the mathematical theorems, methodological rules and the physical postulates constitute a *constraint space*; and consider that scientific theories and their models must be embedded into this space. Einstein was one of the first physicists to become fully aware of the power of constraints, operating as restrictive conditions on scientific constructs. His emphasis on theories of principles, like the theory of relativity, was particularly helpful in this respect. Although Einstein himself did not always clearly distinguish between them, from the modern point of view he imposes four constraints on physical constructs in the theory of relativity. Any *admissible* theory must satisfy such constraints.

Empirical constraints. These constraints comprise Einstein's postulation of the constancy of 'c' in vacuum and his famous predictions: the red shift of light as a function of gravitational field strengths and the bending of light rays in the vicinity of strong gravitational fields. His GTR also explains the perihelion advance of Mercury and other planets.

Principles of Relativity. Einstein characterizes reference frames as 'mechanical scaffolds' or grids, according to which the spatio-temporal

location of bodies can be determined. (Einstein/Infeld 1938, 156) No reference frame must serve as a preferred basis for the description of natural events in the STR. For this reason Einstein abandoned Newton's absolute space and time and 19th century ether theories. Even his Special theory gave an unjustifiable preference to inertial systems and Euclidean geometry. The General theory extends the principle of relativity to all – inertial and non-inertial – coordinates systems. In its general form the principle states that all coordinate systems, which represent physical systems in motion with respect to each other, must be equivalent from the physical point of view. In other words, the laws which govern the changes that happen to physical systems in motion with respect to each other are independent of the particular coordinate system, to which these changes are referred. (Einstein 1905)

Invariance and Symmetry. Invariance is related to the symmetry principles of the relativity theory. In the STR symmetries result from the operations of transformation rules between inertial frames. Reference frames serve as idealized physical systems in the theory of relativity. Compared with the many types of symmetries, which are recognized today (global, local, external, internal, continuous and discrete symmetries, see Castellani 2003, Ch. 26.6; Kosso 2000; Brading/Brown 2004), Einstein only deals with space-time symmetries of a global (STR) or local (GTR) kind. The Lorentz transformations deal with space-time transformations of a global kind: they are constant throughout space and time. The Lorentz transformations represent transformations of the inertial frames, say a boost from a system at rest to an inertially moving system. Symmetry transformations form symmetry groups. Symmetry groups (like the Lorentz transformations) require the physical equivalence of various inertial systems: as we subject inertial frames to transformations (rotation and translation in space-time) certain features remain invariant, others change. The symmetry operations show which physical features remain invariant under the operation of transformations and which features change in the transition between reference frames. As we shall see in the next section many physicists regarded what remains invariant under symmetry operations as the 'real'. According to Einstein only space-time coincidences

can claim physical reality. It seems that the 'invariant' provides a new criterion of what physicists are to count as physically real. A consideration of symmetries also point to the importance of structures in the theory of relativity.

Covariance. The relativity principles state that all inertial and non-inertial systems are to be treated as equivalent from a *physical* point of view. The invariance principle states that symmetry transformations performed, say, on inertial frames must return some values of parameters as invariant. Einstein introduces covariance as 'form invariance' of the laws of physics. (Einstein 1916; Einstein 1922) The laws must retain their form whether they are considered from different coordinate systems or described in different mathematical languages. This intuition reflects Einstein's demand that the laws of physics remain 'covariant' with respect to different coordinate systems of the theory of relativity. We can express the laws of nature in different mathematical languages, for instance in the form of Euclidean or Riemannian geometry. Lorentz covariance means that the form of physical laws must remain invariant as reference systems undergo symmetry operations with respect to their spatial and temporal coordinates. But the covariance constraint takes on its true importance in the GTR. The space-time coordinates are abstract notations, x_1, x_2, x_3, x_4, and the space-time laws are required to remain unchanged under the quite general transformations of the space-time coordinates, which the GTR allows. (Norton 1993, 794-5) Einstein often illustrates covariance with respect to the space-time interval ds^2. (Einstein 1922, 28) In Minkowski space-time, the space-time interval ds^2 is expressed as an invariant expression in what remains essentially a quasi-Euclidean space; for the propagation of light it is:

$$ds^2 = \sum_{\nu=1}^{3} (\Delta x_\nu)^2 - c^2 \Delta t^2 = 0 \quad (1).$$

If the expression satisfies covariance it must remain form-invariant under the substitution of a primed coordinate system, i.e. $ds^2 = 0 = ds'^2$:

$$ds^2 = \sum_{\nu=1}^{3} (\Delta x_\nu)^2 - c^2 \Delta t^2 = 0 = d's^2 = \sum_{\nu=1}^{3} (\Delta x'_\nu)^2 - c^2 \Delta t'^2 \quad (2).$$

The equation for the space-time interval, ds^2, remains form-invariant if *K* is substituted by another quasi-Euclidean inertial frame, *K'*, as indicated by the coordinates $\Delta x_v'$.

The space-time interval, ds^2, is expressed, in Minkowski space-time, by the invariant line element:

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \quad (3)$$

Equation (3) captures Einstein's desire to call his theory 'theory of invariants' rather than 'relativity theory'. Laws must remain covariant under arbitrary transformations of the coordinate systems. It is not easy to say what form invariance actually means. For present purposes it suffices to say that such a change in symbolic form should not affect the objective relations, which the laws encode. Covariance expresses the requirement that equivalent expressions of the laws of nature must remain objective. (Cf. Weinert 2007a) The requirement that the physical laws in the STR and GTR must remain 'form-invariant' under the transformation of space-time coordinates is a further hint, to be developed later, that structures play a significant part in the theory of relativity.

The reason for the imposition of the constraints is to increase the fit between the theory and the world of experience. If the number of constraints and their interconnections can be increased, then many scientific theories will fail to satisfy the constraints. (Einstein 1933, 272; Einstein 1936, 18-9; Einstein 1944, 258) This process of elimination will usually leave us with only one plausible survivor. The General theory of relativity was able to explain the perihelion advance of Mercury, where both classical mechanics and the STR had failed. It would be exaggerated to claim that there exists such a tight fit between the theory and the world that there is a one-to-one mapping of the theoretical with the empirical elements. Due to the need for approximations and idealizations in the theoretical constructs, which are 'free inventions', there will always be theoretical structure, for which there is no direct empirical evidence. But Einstein holds that one theory always satisfies the constraints better than its rivals. It does not follow from this argument that the survivor – let us say the theory of relativity – will be true. It does follow that the process of elimination will leave us with the most

adequate theoretical account presently available. New experimental or observational evidence may force us to abandon this survivor. The desire for unification, logical simplicity and the clash with experience may persuade us to develop alternative theoretical accounts. Einstein's extension of the principle of relativity from its restriction to inertial reference frames in the Special theory to general coordinate systems in the General theory is a case in point. Although Einstein does claim that there is one correct theory, he cannot mean this in an absolute sense. (Einstein 1918b, 226) His insistence on the eternal revisability of scientific theories speaks against this interpretation. What he must mean is that there is always one theory, at any one point in time, which better fits the available constraints. This one theory settles better into the constraint space, which theory and evidence erect, than its rivals. Clearly, Einstein regarded the theory of relativity as a superior theory at his time; proponents of this theory also claimed that it committed them to an invariance view of reality.

Three Views of Reality

1. The laws of physics must express the invariant features, which remain as coordinate systems undergo space-time transformations. Einstein explicitly claims that the laws of physics are statements about space-time coincidences. In fact only such statements can 'claim physical existence'. (Einstein 1918a, 241; Einstein 1920, 95) As a material point moves through space-time its trajectory is marked by a large number of co-ordinate values x_1, x_2, x_3, x_4. The requirement of covariance allows it to be equally well described in terms of the primed coordinates x'_1, x'_2, x'_3, x'_4. This is true of any material point in motion. It is only where the space-time coordinates of the systems coincide that they 'have a particular system of coordinate values x_1, x_2, x_3, x_4 in common'. (Einstein 1916, 86; Einstein 1920, 95) In terms of observers, attached to different coordinate systems, it is at such points of intersection that they can agree on the temporal and spatial measurements of the respective systems. This is Einstein's point-coincidence argument. From this argument, many physicists, including Einstein, concluded as a philosophical consequence of the symmetries of the relativity theory that

only the invariant can be regarded as the physically real. (Einstein 1920, Appendix II) This is now a common-place view:

> All the appearances are accounted for if the real object is four-dimensional, and the observers are merely measuring different three-dimensional appearances and sections; and it seems impossible to doubt that this is the true explanation. (Eddington 1920, 181)
>
> (…) the objective features of the world must be represented by invariant quantities. Why? Because frame-dependent quantities 'change from reference to reference frame' and are, in part, artefacts of convention. (Maudlin 2002, 34)
>
> If two frames from which the universe can be accurately described disagree on a certain matter, then that matter cannot be an objective fact. (Lange 2002, 207; cf. Belashov 1999, 2000)

Yet as Nozick (2001, 329 Fn 11) correctly points out, while frame-specific temporal and spatial measurements in the Special theory of relativity are not invariant but perspectival, they are objective. What effect does this concession have on the invariance view of reality?

2. *Perspectival Reality*. Is it true that only the 'invariant is real'? What happens, say in the STR, to the clock and meter readings in particular inertial frames? As they differ from frame to frame, should we conclude that these events are 'unreal' in the respective inertial frames? Note that the question of the reality or unreality of events in space-time does not depend on observers' perceptual relativity. Different systems in motion with respect to each other register different values for rod lengths and clock times. These measurements do not depend on what observers perceive; they depend on the behaviour of physical systems in motion. For measuring observers in the respective systems, these measurements have *perspectival* reality. Observers in time-like related frames, moving at a constant velocity with respect to each other, can observe that their respective clocks ticks at different rates and their measuring rods do not measure the same lengths. The ticking rate of the clocks and the behaviour of measuring rods show that perspectivalism is not observer-dependent but frame-dependent. It depends on the behaviour

of rods and clocks in particular frames. Only the reading and comparison of clocks depends on the presence of conscious observers. The perspectival realities of physics are the result of a combination of frame-dependent features – the '3+1' view of observers, due to their perspectival lamination of space-time - and frame-independent parameters of inertial frames - the invariant features of four-dimensional Minkowski space-time.

If we adopt perspectival realities, what becomes of the physicist's criterion that *only* the invariant is to be regarded as real? The adoption of perspectival, frame-dependent realities enhances the invariance criterion of reality. The Minkowski space-time structure has both invariant and perspectival aspects. In Minkowski space-time, the non-tilting light cones, emanating from every space-time event, are invariant for every observer. The space-time interval, ds^2, is invariant across inertially moving frames. The particular perspectives then result from attaching clocks and rods to the 'scaffolds'. That is, they result from the particular 'slicing' of space-time by the world lines of inertial systems in relative, constant motions with respect to each other. The space-time symmetries tell us what is invariant across inertial frames, and what is perspectival. Once we know what features remain invariant across different inertial frames, we can derive the perspectival aspects, which attach to different inertial frames, as a function of velocity. Such a modified view of physical reality can be derived from the Minkowski presentation of the theory of relativity. Max Born compared the perspectival realities to projections, which must be connected by transformation rules to determine what remains invariant. The projections are reflections of frame-dependent properties. But there are also frame-independent properties, which are invariant in a number of 'equivalent systems of reference'.

> In every physical theory there is a rule which connects projections of the same object on different systems of reference, called a law of transformation, and all these transformations have the property of forming a group, i.e. the sequence of two consecutive transformations is a transformation of the same kind. Invariants are quantities having the same

> value for any system of reference, hence they are independent of the transformations. (Born 1953, 144)

The Lorentz transformations show, Born adds, that perspectival quantities 'like distances in rigid systems, time intervals shown by clocks in different positions, masses of bodies, are now found to be projections, components of invariant quantities not directly accessible.' (Born 1953, 144)

The theory of relativity leads to the invariance view of reality, which can be modified by incorporating perspectival realities. But Einstein rejected such perspectival views, as they appear in the Copenhagen interpretation of quantum mechanics (QM). In his opposition to the Copenhagen view he appears to adopt a much more traditional view of reality.

In his opposition to the Copenhagen interpretation of quantum mechanics, Einstein is committed to a complete, direct description of reality. (Einstein 1940, 924) By this he means a direct representation of the actual space-time events, rather than a probability distribution of possible outcomes of measurements. Such a complete description of actual events in space-time will avoid non-local effects, the spooky action-at-a-distance, which Einstein found objectionable in QM. For it will be subject to the 'strict laws for temporal dependence.' (Einstein 1940, 923; Einstein 1948, 323; Einstein 1949a, 86) In physics the 'strict laws for temporal dependence' are typically expressed in differential equations, which trace the evolution of some parameter as a function of time. A complete description of quantum reality would recover the differential equations, which describe the temporal evolution of real physical systems in space-time. The Schrödinger equation is of course a differential equation, which spells out the temporal evolution of quantum systems. However, this does not satisfy Einstein, because the Schrödinger equation describes temporal evolution in an abstract Hilbert space. His opposition to the Copenhagen interpretation of QM led him to a more classical *separability view of reality*: spatially separated system, A and B, which obey Einstein locality, possess physical properties, which are not immediately affected by external

influences on either of the systems. (Einstein 1948) This view of reality also transpires in the much-quoted definition of reality in the EPR paper.

If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. (Einstein, Podolski, Rosen 1935, 777; cf. Einsein 1949a, 82-6)

The Importance of Structure

If the models of the theory of relativity *represent* both invariant and perspectival aspects of reality, the question of realism imposes itself. It is generally agreed that Einstein's position shifted from an early sympathy for Mach's positivism to a later commitment to realism. (Einstein 1949a, 10; Holton 1965; Fine 1986; Scheibe 1992, 119; Scheibe 2006, 167-61; but see Howard 1990; 1993) In his 'Autobiographical Notes' he criticizes Mach for having misunderstood the 'essentially constructive and speculative nature of scientific thought' (Einstein 1949a, 20). But the question is which kind of realism the theory of relativity supports.

The *invariance view of reality*, which is a consequence of the introduction of symmetries in the STR, is in good agreement with a certain version of realism, which is expressed in many of Einstein's philosophical announcements. This position simply regards scientific theories as hypothetical constructs, free inventions of the human mind. But science is committed to the existence of an external world, irrespective of human awareness. To be scientific, theories are required to represent reality via models. This version of realism need not claim that the theories, its models and laws are true mirror reflections of the natural world and its regularities. Einstein rejected 'naïve realism'. (Einstein 1944, 280) There only needs to be the objectivity assumption that the models and laws of physics are good approximations and idealizations of the systems modelled. (See Einstein 1949a, 21-2) The models of the Special theory of relativity are idealized representations of kinematic aspects of physical systems. The models of the theory represent specific aspects of the physical systems modelled in the theory.

To focus on the representational aspects of models it will be convenient to distinguish between the topologic and algebraic structure of models. In the simplest case, a model represents the topologic structure of a system; e.g. a heliocentric scale model of the solar system represents the spatial arrangement of the planets around the sun. The models used in the theory of relativity are more sophisticated structural models, which combine a topologic with an algebraic structure. The algebraic structure of the model expresses the mathematical relations between the components of the model. (Weinert 1999; Weinert 2006)

An analysis of the theories of relativity clearly shows that physics is concerned with physical systems, which are modelled in the STR by inertial reference frames and more general coordinate systems in the GTR. The reference frames, characterized by Einstein as 'mechanical scaffolds', select structural aspects of the systems modelled by way of their coordinates; in the STR these are kinematic relations between reference frames in inertial motion. Einstein emphasized his belief in the structure of the real world (e.g. relata and relations) in a number of places:

> 'Without the belief that it is possible to grasp the reality with our theoretical constructions, without the belief in the inner harmony of our world, there could be no science.' (Einstein/Infeld 1938, 296)
>
> 'Physics is the attempt at the conceptual construction of a model of the *real world*, as well as its lawful structure.' (Quoted in Fine 1986, 97; italics in original; Einstein 1948, 321)
>
> The greatest change in the axiomatic basis of physics – in other words, of our conception of the structure of reality – since Newton laid the foundation of theoretical physics was brought about by Faraday's and Maxwell's work on electromagnetic phenomena. (Einstein 1931, 266)

There is clearly a concern with *structure* in Einstein's physics, which is highlighted by the use of coordinate systems as models of reality. The concern with structure is further emphasized by a consideration of Einstein views on structure laws. According to this view the equations of the theory of relativity and electrodynamics can be characterized as *structure laws*, which apply to fields. (Einstein/Infeld 1938, 236-45) Structure laws express

the changes which happen to electromagnetic and gravitational fields. These structure laws are local in the sense that they exclude action-at-a-distance. 'They connect events, which happen now and here with events which will happen a little later in the immediate vicinity.' (Einstein/Infeld 1938, 236) The Maxwell equations determine mathematical correlations between events in the electromagnetic field; the gravitational equations specify mathematical correlations between points in the gravitational field. The postulates of quantum mechanics, like the Born rule, encode the probability of quantum events. Einstein submits that structure laws have the form 'required of all physical laws.' (Einstein/Infeld 1938, 238, 243) According to such a structural view of laws, the laws of physics capture structural aspects of natural systems. That is, they symbolically express the structure of a class of natural systems by showing how their relata are mathematically related to each other. Wigner was similarly aware of the importance of structure 'in the events around us,

> that is correlations between the events of which we take cognizance. It is this structure, these correlations, which science wishes to discover, or at least the precise and sharply defined correlations'. (Wigner 1967, 28; cf. Weinert 2007a)

By associating correlations with structure, Wigner emphasizes that the correlations between events can be mathematically determined; it is the mathematical determination, which provides the structure of the correlation. Generalizing the Einstein-Infeld-Wigner view we can therefore say that structure laws govern how the components (or relata) of physical systems modelled in the theory are mathematically related to each other.

Einstein clearly believes in the existence of a lawlike, structured reality, a physical world consisting of a network of systems, which can be described and explained by physical theories. The constructs of physical theories (axioms, constraints, coordinate systems, laws, models, theorems) express the structure of natural systems in mathematical form. The laws of physics determine the relations between the relata: for instance whether the relations are linear or non-linear, whether they involve quadratic or

polynomial functions. Einstein seems to have believed in the reality of classical objects, fields and the structure of space-time, insofar as it is determined by the matter-energy contents of the universe. Apart from space-time events, the relata may refer to objects like planets (as in Kepler's laws), electromagnetic or gravitational fields or to properties of quantum systems in the wave function, ψ. Einstein declares that 'the concepts of physics refer to a real external world, i. e. ideas are posited of things that claim a 'real existence' independent of the perceiving subject (bodies, fields etc.)' (Einstein 1948, 321, transl. Howard 1993, 238; Einstein 1944, 290)

If the natural systems in the physical world display various kinds of structure, then the models of scientific theories must represent this structure through their algebraic and topologic structures. The mathematical representation of three-dimensional Euclidean space, for instance, takes the form $\langle \Re^3, d \rangle$, where $\Re^3$ represents the Euclidean coordinate systems and d is the Pythagorean distance function. Space-time models can be represented in the general form $\langle M, O_i, CS \rangle$, where M represents the differentiable manifold of space-time points – the topology of space-time points in local neighbourhoods – and the O_i's various geometric objects, like spatio-temporal metrics and the CS represent the coordinate systems of the theory of relativity. The STR is represented by the mathematical structure $\langle \Re^4, h_{ik} \rangle$, where $h_{ik} = diag(-1,-1,-1,+1)$, which is the matrix of the line element $ds^2 = -dx_1^2 - dx_2^2 - dx_3^2 + dx_4^2$. (See Norton 1992, 283, 289; Scheibe 2006, 110-12; Smolin 2006, 205-7)

Such issues of representation suggest that a consideration of the STR naturally leads to some version of structural realism, which predates current debates in the philosophy of science. Structural theories encourage structural explanations since they encourage questions like 'what is the structure of the world like if certain principles are to hold in it?' (Hagar 2008) Structural realism is a thesis about (knowledge of) structural relations. But such relations must, according to Einstein's principle theories, obey constraints. Symmetries constitute one type of constraint. A consideration of the symmetries involved in the STR therefore suggests that the mathematical

relations between relata must include the geometric symmetries, which are important in the STR. Symmetries constitute important elements of structure.

Symmetries and Structure[1]

Physical systems can be regarded as manifestations of structures. A physical structure consists of relata and relations. But the modified invariance view of reality tells us that structures can have both frame-specific and frame-invariant features. What holds the relata together and binds them into specific structures are the mathematical relations. A system, like the solar system, consists of relata (the planetary bodies and the sun) and relations (Kepler's or Newton's laws). The relations prescribe the elliptical orbits of the relata. The job of science is to model mathematical structures, also consisting of relations and relata, which represent the physical structures in an approximate and idealized form. Einstein, for instance, wrote that coordinate frames are modelled as 'representatives' of rigid bodies in mechanics. (Einstein 1925, 538) But Einstein also recognized that the algebraic relations are subject to constraints. Consider, for instance, the effect of symmetries.

Physics is interested in frame-invariant realities, because the frame-specific realities can be obtained from them by the transformation rules of a particular theory. But it seems that all our experience of reality is perspectival or frame-dependent, because of the '3+1' slicing of space-time by observers. In our efforts to obtain frame-invariant realities, symmetries play an important part. Symmetries result from the application of transformation groups, which will leave all frame-invariant parameters unchanged. Frame-invariant parameters are those, which prove to be immune to the possible changes expressed in symmetry operations, like translation in time and space, rotation and mirror imaging. Given appropriate constraints, structures tend to be frame-invariant. But structure also has frame-variant perspectival manifestations. The relations between objects may not be structure-preserving, as for instance the distance relation

[1] This section summarizes the results of Weinert (2007b).

between two objects may change; but the Euclidean distance, r^2, remains invariant.

A transformation group satisfies 3 logical criteria: reflexivity, symmetry, and transitivity. Symmetries can be distinguished according to different properties.[2] Rotation, reflection, spatial and temporal translations and space-time symmetries are typical *geometric* symmetries. They take events, things and properties as their objects. A better name may be *external* (global) symmetries. External symmetries result from the operation of space-time transformation groups. They are external to their reference objects because they govern the invariance of their objects with respect to an 'external' change of space-time reference systems. So according to the Galileo transformations, an event that happens at a location x, can equally be transferred to a location x', because the two locations are related by the equation $x' = x - vt$. Whatever event takes place at location x, its physical structure, expressed in the laws of physics, will not change as a result of its transport to x'.

The characterization of structure as 'relata & relations' in philosophy of science debates cannot be restricted to geometric or physical relations between events and objects alone. The relations themselves are subject to symmetry constraints. The invariance of the relata (fields, objects, properties) are governed by their algebraic relations (laws of nature, symmetries principles, mathematical theorems), which makes the relations structural principles. For the relata to be governed by the relations means for the relations to put constraints on the relata - structural constraints since they determine the type of relata, which are allowed to enter the relations. But the space-time relations themselves are governed by space-time symmetries. Amongst the relations, we find for instance conservation laws, and these follow from symmetry principles, according to Noether's theorem.[3] With respect to the relata, the symmetries are higher-order

---

[2] Wigner (1967); Morrison, (1995); Rosen (1995), 72-6; Earman (1989), 173; Mainzer (1996), 277, 341f, 357, 414, 420 calls dynamic symmetries 'gauge groups'; Cao (1997), Ch. 9

[3] According to Noether's theorem symmetries and conservation laws are related. The laws of conservation are consequences of the space-time symmetry operations. Conservation

principles. If the symmetries preserve invariant parts of structure, then they are constraints on structure. The structure consists of relations and relata, so that symmetries are higher-order constraints on relata and relations. As higher-order principles the symmetries give us the invariant structures, in which physics is interested. Following Leibniz, we can adopt a *relational* approach to structure. Such an approach emphasises that structure is born of a union of relata and relations; relations and relata are equally important for the formation of a structure. We have a triad of relata, relations and higher order principles, such as symmetries. The algebraic relations and symmetries act as constraints on the admissible relations and relata, with the result that, if the constraints do their job, the relata and relations refer to the components of physical structures, albeit in an approximate and idealized way.

An analysis of the STR shows that its consideration in terms of structural realism has to take into account the role of relata - reference frames and coordinate systems – and relations – laws and symmetry principles. *Perspectivalism and invariance are two faces of symmetries.* Symmetries offer a deeper insight into the nature of reality, because they automatically yield frame-specific (perspectival) and frame-invariant properties of physical reality. A suitably modified invariance view of reality may be regarded as support for some ontic version of structural realism, which, however, assumes (against the standard ontic version) the existence of structured physical systems, e.g. the reality of relata and relations, which models aim to represent. This representation is not a plea for naïve realism or even isomorphism between theoretical structures and empirical substructures. The strength of structural realism resides in an awareness of the approximations and idealizations necessarily built into modelling. The only claim made in structural realism, as derived from a consideration of the theory of relativity, is that a certain 'fit' must exists between the theoretical

---

of momentum results from invariance with respect to spatial translations, conservation of energy from invariance with respect to temporal translation, conservation of angular momentum from invariance with respect to spatial rotation and conservation of the centre of mass from invariance with respect to uniform motions. See Mainzer (1996), 350; Feynman (1997), 29-30; Rosen (1995), 150-3, Wigner (1967), 18ff

structure whose job is to represent a physical structure. (Weinert 2006) This fit is secured, as we have seen, through the employment of constraints.

According to the relational view of structure, relata and relations interrelate in such a way that it can be misleading to claim that structures are prior to relata. Events, objects, properties and systems are needed as inputs to structures. Symmetries help to determine the invariant parts of structures. But invariant structures also need the input of relata.

Conclusion.

Looking at Einstein's constructive work and some of his diverse statements on realism in science, we have argued that the STR commits its proponents to a certain form of structural realism. Such a commitment is implicit in the representational claims associated with Einstein's focus on principle theories, their models and constraints. The invariance view, suitably modified to include perspectivalism, anticipates the emphasis on structure, which has dominated the recent debate about Structural Realism. In particular the invariance view highlights the role of space-time symmetries, which result in invariant and perspectival aspects of the systems, to which the transformation groups are applied. Symmetries also constitute an important form of constraint. An analysis of the Special theory of relativity tells philosophers much about science, which has not been sufficiently analyzed in the literature, in particular, the central role of *constraints* in scientific theorizing and modelling.

References

Balashov, Y., 1999, Relativistic Objects, *Noûs* 33, 644-62.

Balashov, Y., 2000, Relativity and Persistence, *Philosophy of Science* 67, 549-62.

Born, M., 1953, Physical Reality, *The Philosophical Quarterly* 3, 139-49.

Brading, K./E. Castellani, *eds*., 2003, *Symmetries in Physics*, Cambridge University Press.

Brading, K./H. R. Brown, 2004, Are Gauge Symmetry Transformations Observable? *Brit.J.Phil.Sci.* 55, 645-65.

Cao, T. Y., 1997, *Conceptual Development of Quantum 20th Field* Theories, Cambridge University Press.

Castellini, E., 2003, On the meaning of symmetry breaking, in Brading/Castellani, *eds*., 2003, 321-34.

Earman, J., 1989, *World Enough and Space-Time*. MIT Press, Cambridge (Mass.)/London.

Eddington, A. S., 1920, *Space, Time and Gravitation*, Cambridge University Press.

Einstein, A.,1905, Zur Elektrodynamik bewegter Körper, *Annalen der Physik* 17; reprinted in H. A. Lorentz/A. Einstein/H. Minkowski, *Das Relativitätsprinzip*, Wissenschaftliche Buchgesellschaft, [7]1974, 26-50, Darmstadt.

Einstein, A., 1916, Die Grundlage der allgemeinen Relativitätstheorie, reprinted in H. A.Lorentz/A. Einstein/H. Minkowski, *Das Relativitätsprinzip*, Wissenschaftliche Buchgesellschaft, [7]1974, 81-124, Darmstadt.

Einstein, A, 1918a, Prinzipielles zur allgemeinen Relativitätstheorie, *Annalen der Physik* 55, 241-44.

Einstein, A, 1918b, Prinzipien der Forschung, reprinted in Einstein, 1977, 107-10; English translation in Einstein, 1954, 224-7.

Einstein, A., 1919, Was ist Relativitätstheorie?, reprinted in Einstein, 1977, 127-131; English translation in Einstein, 1954, 227-32.

Einstein, A., 1920, *Relativity* - The Special and the General Theory, Methuen, London.

Einstein, A., 1922/[15]1954, *The Meaning of Relativity*, Methuen, London.

Einstein, A., 1925, Die Relativitätstheorie, in *Collected Papers*, Volume 4 (1912-1914), in M. J. Klein, A. J. Kox, J. Renn, R. Schulmann, *eds*., Princeton University Press 1995, 536-50.

Einstein, A., 1931, Maxwells Einfluss auf die Entwicklung der Auffassung des Physikalisch-Realen, reprinted in Einstein, 1977, 159-162; English translation in Einstein, 1954, 266-70.

Einstein, A., 1933, Zur Methodik der theoretischen Physik, reprinted in Einstein, 1977, 113-119. English translation in Einstein, 1954, 270-76.

Einstein, A., 1936, Physics and Reality, *Journal of the Franklin Institute* 221, 348-382; reprinted in Einstein, 1954, 290-323.

Einstein, A., 1940, Considerations Concerning the Fundaments of Theoretical Physics, *Nature* 145, 920-4.

Einstein, A., 1944, Remarks on Bertrand Russell's Theory of Knowledge; quoted from *The Philosophy of Bertrand Russell* in A. Schilpp, *ed*., New York: Tudor Publishing Company, [3]1951, 278-91.

Einstein, A., 1948, Quantenmechanik und Wirklichkeit, *Dialectica* 2, 320-4.

Einstein, A., 1949a, Autobiographisches/Autobiographical Notes, in Schilpp *ed*., 1949, Volume I, 2-94.

Einstein, A., 1949b, Replies to Criticisms, in Schilpp *ed*., 1949, Volume II, 665-88.

Einstein, A., 1950, On the Generalized Theory of Gravitation, *Scientific American* 182, 341-56.

Einstein, A. 1954, *Ideas and Opinions*, London: Alvin Redman.

Einstein A. 1977, *Mein Weltbild*. Hrsg. von Carl von Seelig, Ullstein, Frankfurt a./M.-Berlin-Wien.

Einstein, A./B. Podolsky/N. Rosen, 1935, Can Quantum-Mechanical Description of Physical Reality be Considered Complete?, *Phys.Rev.* 47, 777

Einstein, A./L. Infeld 1938, *The Evolution of Physics*, Cambridge University Press.

Feynman, R. P., 1997, *Six Not So Easy Pieces*, Helix Books, Perseus Books, Cambridge (Mass.).

Fine, A. 1986, *The Shaky Game* - Einstein, Realism and the Quantum Theory, The University of Chicago Press, Chicago and London.

Holton, G., 1965, The Metaphor of Space-Time Events in Science, *Eranos Jahrbuch* 34, 33-78.

Hagar, A., 2008, Length Matters, in *Studies in History and Philosophy of Modern Physics* 39, 535-56.

Howard, D., 1990, Einstein and Duhem, *Synthese* 83, 363-84

Howard, D., 1993, Was Einstein Really a Realist?, *Perspectives on Science* 1, 204-51.

Kosso, P., 2000, The Empirical Status of Symmetries in Physics, *Brit.J.Phil.Sci.* 51 81-98.

Lange, M., 2002, *An Introduction to the Philosophy of Physics,* Blackwell, Oxford.

Mainzer, K., 1996, *Symmetries of* Nature, Walter de Gruyter, Berlin/New York.

Maudlin, T., 1994/[2]2002, *Quantum Non-Locality and Relativity*, Blackwell, Malden (Mass.)/Oxford.

Morrison, M., 1995, *Symmetries as Meta-Laws*, in F. Weinert *ed*., 1995, 157-88.

Norton, J. 1992, The Physical Content of General Covariance, *Einstein Studies* 3, 281-315.

Norton, J., 1993, General covariance and the foundations of general relativity: eight decades of dispute, *Rep.Prog.Phys.* 56, 791-858.

Nozick, R., 2001, *Invariances*, The Belknap Press at Harvard University Press, Cambridge (Mass.)

Rosen, J., 1995, *Symmetry in Science*, Springer-Verlag, New York, Berlin, Heidelberg.

Scheibe, E., 1992, *Albert Einstein, Theory, Experience and Reality*, reprinted in E. Scheibe, 2001, *Between Rationalism and Empiricism,* Springer, New York-Berlin-Heidelberg, 119-35.

Schilpp, P. A. *ed*., 1949, *Albert Einstein: Philosopher-Scientist*, 2 Volumes, Open Court, La Salle (Ill.)

Scheibe, E., 2006, *Die Philosophie der Physiker*, C. H. Beck, München.

Smolin, L., 2006, The Case for Background Independence, in: *The Structural Foundations of Quantum Gravity*, D. Rickles, St. French, J. Saatsi *eds*., Clarendon Press Oxford, 196-39.

Weinert, F. *ed*., 1995, *Laws of Nature* – Essays on the Philosophical, Scientific and Historical Dimensions, Walter de Gruyter, Berlin.

Weinert, F., 1999, Theories, Models and Constraints, *Studies in History and Philosophy of Science* 30**,** 303-333.

Weinert, F., 2006, Einstein and the Representation of Reality, *Facta Philosophica* 8, 229-52.

Weinert, F., 2007a, Einstein and the Laws of Physics, *Physics and Philosophy* (Issue 2007), 1-27, http://physphil.tu-dortmund.de/.

Weinert, F., 2007b, Invariance, Symmetries and Structural Realism, *PhilArchive*, http://philsci-archive.pitt.edu/archive/00003643/

Wigner, E., 1967, *Symmetries and Reflections*, Indiana University Press, Bloomington/London.

Vagueness and Paradox
(Ontology at the Limit)

Virgil DRĂGHICI [*]
Babes-Bolyai University Cluj-Napoca

Abstract:

The aim of this paper is to analyse some logic and philosophical aspects of vagueness, e.g. the sources of vagueness, the paracomplete solution(s) to sorites paradox, the existence of higher-order vagueness, the soundness of Evans argument and the vagueness in semantical paradoxes. Afferent, the problems involved by the ontology constructed at the limit of paradoxicality are discussed. All these matters are considerated in a double register: with arguments *pro* and *contra.*

Keywords*:* vagueness, sorites, paracomplete logic, higher-order vagueness, Evans argument, realism, ontology.

Vagueness

Some predicates like "red", "hot", "rich", "bold" are usually taken as vague predicates, that is, they are examples in which no sharp line can be drown separating predicate's positive extension from its negative extension. In other words, a predicate is vague if there are *borderline cases* for it. Though present in the work of prominent contemporary authors[1], this notion of vagueness has deep historical roots. It arose in antiquity in the context of *sorites paradoxes*, some apparently valid arguments based on apparently

[*] E-mail: vidra007@gmail.com.
[1] E.g. Russell (1923), Church (1960), Quine (1960).

true premises and with apparently false conclusion. The standard form of a sorites is a chain of the following form:

(I) $\varphi (a_1)$
$\varphi (a_1) \rightarrow \varphi (a_2)$
$\varphi (a_2) \rightarrow \varphi (a_3)$
…
$\varphi (a_{n-1}) \rightarrow \varphi (a_n)$
_____
$\varphi (a_n)$, with n arbitrar, and φ a soritical predicate ("bald").

An argument is soritical only if the predicate φ appears determinately true of a_1, determinately false of a_n and each pair (a_k, a_{k+1}) in the ordered series $< a_1, \ldots, a_n >$ appears indiscriminable in respect of φ.

But a soritical argument can have other forms. If in (I), for example, the set of conditional premises is replaced by a universally quantified premise, then the sorites has the form of mathematical induction.

(II) $\varphi (a_1)$
$\forall n (\varphi (a_n) \rightarrow \varphi (a_{n+1}))$
_____
$\forall n \varphi (a_n)$

That is, if a man with one hair is bald and the addition of one hair is not relevant for the distinction bald/not bald, then a man with n hairs (regardless of n) is bald.

Or, as in the following argument:

(III) 1. A man with one hair is bald.
2. A man with 10^4 hairs is not bald.
3. So there must be a least such number k such that a man with k hairs is bald and a man with k+1 hairs is not bald, symbolically represented as

$\varphi(a_1)$
not $\varphi(a_n)$

$\exists k\,(\varphi(a_k) \wedge \text{not } \varphi(a_{k+1}))$

This is a variant of (II), based on the least-number principle, equivalent to the principle of mathematical induction.

What is the solution to this kind of paradox?

Is the solution independent of some extra-logical items, e.g. considerations regarding the roots (sources) of vagueness? Where can be set the logic (in a wide sense) in respect of the vague concepts?

By considering firstly this latter question, a variety of positions can be detected.

1. The logic and vagueness (fuziness) of some linguistic expressions have nothing in common.
2. The logic is applicable, but according to classical logic the argument is valid. However, it cannot be accepted, therefore the *classical* logic must be rejected.
3. The logic is applicable to vague expressions, the soritical argument is an example of a *valid* argument, with a false conclusion, hence with one premise false.
4. The logic is also applicable, the argument is valid, the premises are true; therefore the conclusion must be accepted.

Relating to the source of vagueness, four options can be mentioned:

a) The *epistemic view*,[2] according to which the lack of the precise boundaries in the application of a vague predicate is due to our inevitable ignorance. Therefore, the indeterminacy is *merely* apparent; it is only apparent that the extension of the predicate "bald" has no sharp boundaries, yet such boundaries there are, though we do not know where. The epistemic view is a form of a *robust semantic realism*.

---

[2] Represented by Williamson (1994) and Sorensen (1998), (2001).

b) A moderate form of epistemicism, *semantical realism*, according to which we do not know where the bourdaries of the "bald" lie, for in this case *there is no determinate fact of the matter*, and therefore such questions have no determinate answer. We do not know where is the sharp boundary, for such a boundary don´t exist. Vagueness is *merely* semantical, not ontological in any way. This option is compatible with 1, 2 and 3 above.
c) The vagueness is *ontological.* The vagueness of "bald" is due to the vagueness of the corresponding property denoted by this predicate: baldness. This option can adopt a solution of type 2 or 3 to the soritical arguments.
d) The vagueness is *pragmatic*: "[a]ny talk about our vague, natural language should then be reductible to sentences about our (vague) use of precise languages."[3]

Is "vague" vague?

Russellian theory of vagueness is usually taken as the traditional point of view regarding vagueness. A predicate like "bald" is vague, having thus borderline cases. But in spite of the fact that "some men are certainly bald, some are certainly not bald, white between them there are men of whom it is not true to say they must either be bald or not bald"[4], this vagueness of "bald" is not ontological:

> There is a certain tendency in those who have realized that words are vague to infer that things also are vague… This seems to me precisely a case of the fallacy of verbalism - the fallacy that consists in mistaking the properties of words for the properties of things. Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. They have to do with the relation between a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end of it. Nothing is

[3] Keefe (2000), 142.
[4] B. Russell, 1923, 85-86.

more or less what it is, or to a certain extent possessed of the properties which it possesses.[5]

Can this view regarding vagueness affects in some way (classical) logic? Not at all. For the structure of the world is "classical" and can be represented by an *ideal language*, as the deep structure of the natural language, and classical logic is just the logic of this language. Vagueness is a feature of natural language, hence it does not affect in any way the classical logic. To be sure, the two fundamental theses underlying the relationship logic-world-language are: the world is *(a priori)* precise and can be described by an ideal language, and classical logic gives the structure of this language.

The same result regarding the relationship between the vagueness of the natural language and ontology can be attained if we renounce the *a priori* claim of precise character of the world in favour of an eliminativist point of view. According to this view the world is precise, for we can describe it completely in a precise ideal language.[6] This can be obtained by elimination of all vague terms, while preserving its descriptive power; more exactly by the construction of the ideal scientific language, either by precisification of the vague terms or by replacing them with precise terms. Of course, a lot of examples can be given, ilustrating the impossibility of construction of an adequate "image" of the world only by using of such an ideal language. Moreover, the idea of reduction of vague discourse (i.e. a discourse containing vague terms) to a precise one has some "illogical" in it. A term is vague if it has some borderline cases. How, then, can it be reduced, *equivalently* (i.e. co-extensivelly), to a precise one, that is, to a term *without* borderline cases?

---

[5] 84-85.

[6] To be sure, there are some differences between the advocates of this view: Carnap (1950), (1966), for example, claims that all the vague terms can be dispensed with, without any loss of descriptive force of the language, while Quine (1981) sustained that the elimination of the whole class of vague terms does affect the descriptive completeness, but the gain is indisputable: simplicity of the theory and the preservation of classical logic.

The idea of elimination of vague terms can have yet more drastic ontological consequences. According to Unger and Wheeler[7] the vague terms are soritical and therefore incoherent and they must be eliminated, regardless of ontological consequences. Unger (1979)b, for example, considers the following set of propositions: 1. There is at least one stone. 2. If it is a stone, then it consists of many atoms, in a finite number, say *n,* and 3. The removal of one atom does not affect the fact that what remains is a stone. It is easy to argue that this set is inconsistent, for by removing *n* atoms, there are no atoms left at all, but we can suppose that there is a stone (by 3), a fact contradicting 2. So? Unger's solution is sceptical: "(h)owever discomforting it may be, I suggest any adequate response to this contradiction must include a denial of the existence of even a single stone."[8]

The treatement of soritical terms and that of their denotations are similar[9]: if soritical terms are inconsistent, the objects referred to do not exist.

Is Unger's and Wheeler's view coherent? Answering that requires a more subtle analysis, for it implies the problem of higher-order vagueness. That is, it entails an answer to the question: is "vague" vague? If that is the case, then "vague" (i.e. "soritical") is vague, hence *not* inconsistent, and the above view appears to be itself incoherent. Let us see.

Many authors[10] have seriously considered the question whether higher-order vagueness is an essential feature of vagueness as such or not.

Sorensen (1985) has argued that the predicate "vague" is itself vague, in the following way.

Firstly, a vague predicate like "small" generates soritical paradox, for from two seemingly true premises, a) 1 is small and b) For every integer *n*, if *n* is small, then *n+1* is small, a false conclusion follows: c) Therefore, 10^{10} is small. Secondly, a numerical predicate, "*n*-small", can be defined for every integer n, thus:

---

[7] Unger (1979), a,b,c, (1980), Wheeler (1979).
[8] 120 f.
[9] "our results concern words and things alike", 147.
[10] E.g. Sorensen (1985), Wright (1987), (1992), Tye (1990), (1994), Hyde (1994), (2003), (2008), Varzi (2003).

Def: k is n-small if k is either small or less than n.

Clearly, the predicate "1-small" is as vague as "small", both apply to 0 and to any other integers, in the same way. But if $n=10^{10}$, i.e. clearly not small, the extension of the predicate "n-small" is sharply defined only by the second disjunct of *Def;* "less than n": every integer less than 10^{10} is 10^{10}-small. But between the predicates with borderline cases and those without borderline cases there is no clear demarcation line (i.e. there is not a clear value of n making such a difference), a soritical argument for "vague" can be constructed in a similar fashion:

1. "1-small" is vague
2. For every integer n , if "n-small" is vague , then "$n+1$-small" is vague.
3. Therefore "10^{10}-small" is vague.

And, consequently, *the predicate "vague" is vague.*

Is Sorensen's argument sound?

In Tye's view (1994), 44, it is not sound, for the predicate "vague"is not vague, according to the following argument. If "vague" were vague, then there would be vaguely vague predicates (assuming the soundness of Sorensen's argument). But an alternative explanation to the soriticality of "vague" as it appears in Sorensen´s argument can be given. Being vaguely vague entails not being vague. Therefore "vague" is not vague, and the argument is not sound.

Is Tye's argument sound? It would seem that it is not. A very convincing line of reasoning is given in Hyde. In his view Sorensen's argument implies the existence of higher-order vagueness. But, essentially, this idea "is already entailed by the paradigmatic conception and can be seen to follow when the notion of 'border case' employed therein is properly understood" (1994), 39. By using the "paradigmatic conception" (i.e. the idea of vagueness being defined by the existence of borderline cases) the incoherence of Tye's argument can easy be proved.

Firstly, let us see how the paradigmatic conception entails the existence of higher – order vagueness. The Hyde's argument runs as follows:

1. "… is vague" means "there are borderline cases of …" (by paradigmatic conception);

2. The predicate "vague" is vague;

3. Therefore, "there are borderline cases of …" is vague (by 1. and 2.);

4. Hence "borderline case of…" is vague[11];

5. Hence, there are borderline cases of "borderline case of…";

6. Hence, there are predicates that have borderline borderline cases.

Undoubtedly, this argument is sound. The key step in this argument is evidently 2, coupled with the idea of "paradigmatic conception".

Secondly, returning to the Tye's argument, Hyde argues that it is not sound, for an inconsistency can be proved by using the same "paradigmatic conception" (Hyde, (2008), 29): [12]

> […] suppose that the predicate "vague" is vaguely vague. Thus it is indeterminate whether "vague" is vague. Yet this is just equivalent to claiming that "vague" is a borderline case for the predicate "vague". Thus "vague" has a borderline case – namely itself. So "vague" is vague.

In summary, the predicate "vague" is vague, and given that being vague entails *not* being vaguely vague, the conclusion is: "vague" is homological in determinate way, that is it is vague and *not* vaguely vague[13], contradicting thus Tye's view. Notice that what Hyde's argument establishes is a stronger result (Hyde, (2008), 29):

> Not only is "vague'' vague, however. In light of the foregoing reasoning we can also see that *some predicate is vaguely vague if and only if "vague'' is vague*, and therefore recognize the existence of predicates that are vaguely vague. Such predicates present us with examples of borderline cases of borderline cases […]

---

[11] By the idea that if in a expression a part of it is precise then its vagueness is determinated by the other parts. In the "there are borderline cases of…", the part "there are" is precise.

[12] Comp. and Hyde (2003), 302.

[13] "Higher – order vagueness is a real phenomenon. We can neither claim that it determinately does not exist nor that it is vague whether it exists. It determinately exists" (Hyde (2008), 29).

Weakly paracomplete logic and representionalism

A logic is *paracomplete* if it is *incomplete* (i.e. for some sentence α, neither it nor is negation is true, for any valuation, that is α is gappy), and is *non-implosive* (i.e. it rejects the spread–principle: if there are gaps anywhere, then they are everywhere). Paracomplete logics cover a wide variety (e.g. weakly paraconsistent logic[14], Lukasiewicz three-valued logic, Kleene three-valued logic). They are required by a variety of *gap theories:* supervaluation-style gap theories (gap theories based on supervaluational fixed points, those based on revision-rule constructions, axiomatic theories[15]) or Kleene-like gap theories (e.g. the theory FM).

Supervaluationism is a theory considered in its applications to vagueness and also to Liar-type constructions. For the first task the standard reference for it is Fine (1975).[16] For the supervaluationist the truth *simpliciter* is the *determinate* truth or *supertruth.* For the concept of truth as such bivalence fails and for the borderline cases of a predicate there are corresponding truth-value gaps: *φ(a)* is neither true not false (it is indeterminate) if *a* is a borderline case. For the precise predicates the supervaluationist semantic is classical. This kind of supervaluationism is *classical supervaluationism*, its non-bivalent semantics being an extension of the classical one. But what about the laws of classical logic? On this account the conjunction is non-truthfunctional: sometimes it is indeterminate (if α is indeterminate, then $\alpha \wedge \alpha$ is equivalent to α, also indeterminate) and sometimes determinate ($\alpha \wedge \neg \alpha$ is false, α and $\neg \alpha$ are contradictory); similarly for disjunction, The law *tertium non datur* is thus preserved, for $\alpha \vee \neg \alpha$ is true even if α and $\neg \alpha$ are both indeterminate. The relation of classical consequence is valid, are also valid the principles of conditional proof, proof by cases, *reductio*, contraposition, *modus ponens*.[17]

Perhaps the problematic part of classical supervaluationism is the meaning of disjunction, for it entails the validity of *tertium non datur* and,

---

[14] Cf. Arruda (1989).
[15] System H of Friedman and Sheard (1987), System VF of Cantini (1990).
[16] Fine (1975).
[17] This is a weakly paracomplete logic, required by classical supervaluationism. In a language with a determinacy operator, D, all these principles fail.

apparently, the precision of every predicate considered. Apparently, for in the supervaluationist gap theory from "$\alpha \vee \neg \alpha$" is true does not follow "α" is true or "¬α" is true, given the non-bivalent semantics of this theory. But a problem remains, that of the counterintuitive acceptance of *tertium* in treating the phenomenon of vagueness. What is the philosophical view that tolerates consistently both ideas: the validity of *tertium* and the preserving of vagueness? According to Fine (1975) this would be representationalism, that is, the view according to which the roots of vagueness are *merely* semantic, and, therefore, if we cannot explain precisely the world, then we cannot say how it is, hence this idea is not inconsistent with the acceptance of *tertium non datur*.

This is also the view of Dummett (1975), 311:

> if we suppose that all vagueness has its source in the vagueness of certain primitive predicates, relational expressions and quantifiers, we may stipulate that a statement, atomic or complex, will be definitely true just in case it is true under every sharpening of the vague expressions of kinds which it contains.

Considerations parallel to those regarding the non-distributivity of "true" over disjunction can be made in respect of the validity of sorites. The supervaluationist takes the inference in a sorties "bald"-like as valid but unsound, for its major premise (the quantified) is false. Therefore, being fals that "for any *n*, if a man with *n* hairs is bald, then a man with *n+1* hairs is bald", it results that it is true that there is some *n*, such that a man with *n* hairs is bald whilst a man with *n+1* hairs is not bald. And the true of the last claim would be equivalent to the existence of a sharp line between "bald" and "not bald", that is, the problem of vagueness of "bald" just disappears. To avoid such a consequence the distributivity of "true" over "there is" must be denied. That is, a) *True "*$\exists n$ *(Bald (n)* $\wedge \neg$ *Bald (n+1)"* does not imply b) $\exists n$ *True "(Bald (n)* $\wedge \neg$ *Bald (n+1))."* Cumbesome? Not at all, according to Fine's "truth-value shift", a) is true, and its acceptance does not imply the erase of vagueness of "bald"; by contrast, b) would imply. What, more exactly, means a)? Two apparently incompatible ideas can be

contained in a): the existence of a sharp line for "baldness" and the vagueness of it. Again, a representationalist approach to vagueness can "resolve" the puzzle: the predicate "bald" is vague, but the property it denotes is precise, even if we cannot say *where* exactly is the limit. Fine's "truth-value shift" is just the idea of the limit shift among the different precisifications of the vague predicate "bald". More exactly, we suppose that there are three persons a_1, a_2, a_3 and that a_1 is determinately tall, a_3 is determinately not tall and a_2 is a borderline case for "tall". Therefore, "tall" is vague, for there are two different ways to make it precise, or to precisify it, by including or not in its extension the person a_2. Then "truth-value shift", with reference to our predicate "bald" means that in every precisification there is a hair-splitting *n* (the case a), and *not* that there is an *n* it limits the predicate "bald" in every precisification.

Undoubtly, the meaning of "there is" in this supervaluationist approach is *non-standard*.[18] Its sense, in this theory, seems to be consistent. And the metaphysical support of supervaluationist seems to be the representational one.

But can the representationalist view so easy be rejected? Apparently not! For in favor of it speaks a strong argument (Evans). But is also true that a carefully analysis of it shows the problematic ideas underlying this argument.

Evans argument[19] (pro and contra)

In spite of its very short and apparently clear structure the argument has often been misunderstood. It concerns with the possibility of existence of vague objects, that is, with the possibility

> that the world might itself be vague. Rather than vagueness being a deficiency in our mode of describing the world, it would then be a

[18] Comp. and T. Williamson (1994), 153: "According to supervaluationism, "p or q" is sometimes true when no answer to the question "Which" is true. For similar reasons, "Something is F" is sometimes true when no answer to the question, "Which thing is F?" is true. In this sense supertruth is elusive."
[19] Cf. G. Evans (1978).

necessary feature of any true description of it. It is also said that amongst the statements which may not have a determinate truth-value as a result of their vagueness are identity statements. Combining these two views we would arrive at the idea that the world might contain certain objects about which it is a fact that they have fuzzy boundaries. But is this idea coherent?

Firstly, according to Evans, "(c)ombining these two views" means that the existence of vague identity statements is the *necessary* condition for the possibility of there being vague objects. That is, the following conditional holds: if there are not vague identity statements, there are not vague objects. Equivalently, if there are vague objects, then there are vague identity statements.[20]

Secondly, according to Evans view, by *reductio* can be proved that the idea of existence of vague identity statements is inconsistent (Evans proof). Therefore, abstract objects do not exist.

Let us see Evans proof.

A vague identity statement is a statement whose truth-value is indeterminate. If a and b are singular terms, then $\nabla(a = b)$, where "∇" is a sentence operator expressing idea of indeterminacy, is a symbolic representation of a vague identity.

1. $\nabla(a = b)$

2. $\hat{x}[\nabla(x = a)]\ b$, obtained from 1. where "$\hat{x}[\nabla(x = a)]$" is an abstractor denoting the property of "being vague identical to a", a property ascribable to b.

3. $\neg\nabla(a = a)$ as a non-indeterminate fact of self-identity .

4. $\neg\ \hat{x}\ [\nabla(x = a)]\ a$

5. $\neg\ (a = b)$ from 2, 4 and Leibniz's Law: if a and b do not share the same properties, they are different.

---

[20] In the same spirit, Sainsbury (1988), 4 writes: "… if an object were vague, it would be a vague matter what object it is identical with". Garrett (1988), 130 makes even a stronger claim, *Vague identity Thesis*: "The thesis that there can be vague objects is the thesis that there can be identity statements which are indeterminate in truth-value (i.e. neither true nor false) as a result of vagueness […]." In the same vein, Wiggins (1986), 173.

But 5 and 1 are inconsistent: 5 asserts the *falsity* of identity statement, whilst 1 asserts its *indeterminacy.*

Is Evans proof valid, is Evans argument sound? Sometimes the proof was considered invalid, for the inference from 1 to 2 would be fallacious by involving the quantification within the scope of the indeterminacy (∇). To be sure, the validity of this inference will depend on the interpretation of the indeterminacy, *de dicto* or *de re,* i.e. the vagueness will be attributable to the semantic vagueness of the terms involved, or it is attributable to the ontological vagueness. Of course, if the indeterminacy in 1 is merely semantic, then it cannot be attributable to any object. Therefore, the attribution of the property of indeterminacy to *b* is fallacious.[21] Hence, the validity of inference from 1 to 2 requires that the terms *a* and *b* be precise designators, i.e. is not vague what the respective designators denote, for if a designator is imprecise, then no object is determinately denoted by the respective term. Therefore, if 1 is interpreted as *de re* vague identity, then the validity of step form 1 to 2 is guaranteed.

According to some authors[22] even in *a de re* interpretation of indeterminacy, the step from 1 to 2 is problematic for it entails an inconsistency, given the idea that the abstract "$\hat{x}[\nabla(x = a)]$" denotes a property. The inconsistency has as source the *impredicative nature* of the property denoted by the abstract.[23] For according to Leibniz's Law the identity $a = b$ means the coincidence of all properties, that is $\forall\varphi(\varphi(a) \equiv \varphi(b))$, the indeterminacy of the former being equivalent to the indeterminacy of the latter. Therefore, the abstract $\hat{x}[\nabla\ \forall\varphi\ (\varphi(\mathrm{a}) \equiv \varphi(\mathrm{b}))]$ denotes just this idea, a property defined by reference to a totality (of all properties) of which it is part. And this impredicativity is the source of an inconsistency.

---

[21] Comp. Lewis (1988), 128-129. For Sainsbury (1995), 66-68, the proof is valid even for a *de dicto* interpretation of ∇ (a=b) in 1. As a conclusion, there are no vague *names*.

[22] Parsons (2000), Hyde (2008).

[23] A similarity between the above abstract and Russell´s set-abstract "the set of all sets that are not members of themselves", or Grelling´s abstract "the property of all properties that are not applicable to themselves" reveals the impredicativity.

Another argument regarding invalidity of Evans proof concerns the idea of misinterpreting the indeterminacy (in Evans abstract) as a property. According to Keefe (1995), 187-188, the vagueness of $a = b$ in 1 means simply, *there is no fact of the matter as to whether a is b*, from which we cannot derive that this is a property attributable to *b*:

> the key issue for the assessment of [the Evans Proof] is whether delta-predicates (i.e. abstracts involving "∇") denote properties … I maintain that they do not. The indeterminacy operator plays the role of indicating that it is indeterminate whether something has a given property. Expression of this should not be taken to be the (determinate) ascription of another property. If it is indeterminate whether *a* is *F*, there is no fact of the matter about whether it is *F* – the facts do not thereby determine that *a* has a property accounting for this indeterminacy.
>
> Not every statement about *a* can be construed as specifying that *a* has a property. Some statements might describe a mode in which it has a property, whilst … statements containing the indeterminacy operator express that it is indeterminate whether it has a given property. If we allow that it can be genuinely indeterminate whether something has a particular property, then we must deny that "it is indeterminate whether …" denotes a further property. Assuming that it does … begs the question against this possibility.

To be sure, the rejection of the idea that Δ- predicates denotes properties has important *ontological* consequences. Consider, for exemple a cat C with some single hair *h* as a borderline part. Consider then a cat C* that is the sum of C and *h*. Is C= C* a *vague* identity statement or is it determinate? If we agree, as Morreau[24] does, that the Δ- predicates denote properties, then the property "has-h-as-a-determinate-part" will determinately differentiate C* from C. Therefore, the statement C = C* is not vague, it is simply false. But if the Δ- predicates do not denote properties, Hide's view (2008), 143, then the things are different, *de re* vague identity statements must be admitted:

---

[24] Morreau, M (2002).

> In fact, where a *mereological precisification of an object x* is an object *x´* having as determinate parts anything determinately part of *x* and having as determinate non-parts anything determinately not a part of *x*, and having any borderline part of *x* as a determinate part or determinate non-part (so having no borderline parts and being merelogicaly precise), these principles will ensure that it is *de re* vague whether *x= x´*.

In conclusion, Morreau will admit of the existence of vague objects, but will reject the idea of vague identity.[25] Hence, Evans proof is valid, but Evans argument is unsound, for even if the idea of vague identity is inconsistent, it is not a *necessary* condition for the existence of vague objects. By contrast, Hyde endorses a more realistic view, according to which the existence of vague objects entails (with some philosophical conditions) the existence of the *de re* vague identities. However, Evans argument is unsound, given the invalidity of Evans's proof.

Vague objects, vague existence

Now our question is whether there are vague objects, that is, if the vagueness is *ontological.* A simple, common sense reflection on some objects like cats, persons, mountains, living entities, clouds etc cast doubt over the idea of their being precise. A cat, for example, for which there are 50 hairs in the process of coming loose gives a case of indeterminacy: for each hair it is vague whether it is a part of the cat. And this is only apparently a joke, for such a case has generated many views regarding the existence of the cat: there exists a vague cat (a single object spatially vague, with borderline parts, Hyde); there is no vague cat, but many cats and only precise objects must be admitted; there are no vague objects at all (Unger); there are many cats, but they are almost one (Lewis); the many cats are different objects but the same cat (identity is relative, Geach (1980)); some one candidate is the cat, even if it is not true of any candidate that it is the cat (supervaluationism).

---

[25] Similar views: Burgess (1990), Tye (2000), Williamson (1994).

What it is for an object to be vague? From the above example it is clear that the spatial indeterminacy is taken into account[26], that is, some hairs are neither determinately a part of the cat, nor determinately not a part of the cat. But if we admit the existence of vague objects, are we committed to the idea that there are cases of vague existence? A number of different views can be pointed out, e.g. the Lewis-Sider argument, according to which the rejection of ontological vagueness implies the rejection of vague existence[27], Morreau's view (2002), 336, according to which the idea of existence as a matter of degree is counterintuitive[28], Hyde's view, (2008), 137, according to which "[t]he vagueness of the distinction between being and not being is no more mysterious than the vagueness between being a part and not being a part. Sometimes there may simply be no fact of the matter." And, as in this last case, if we accept the idea of vague existence, what sense it is suppose to have? According to Hyde from the idea of vagueness of an object we cannot deduce that there is some object for which it is vague whether it exists. However, the idea of vague existence can be admitted by considering it, via Russell, as a second-order property: "to claim something exists is to claim that some property or other itself has the property of being instantiated. Vague existence then amounts to vagueness as to whether that property itself has the property of being instantiated." (2008), 138.

Therefore, if the Evans argument is not sound and the weakly paracomplete logic associated with representationalism is inadequate as a logic of vagueness and if the ontological vagueness seems to be admitted, then which is that adequate logic of vagueness?

Logic of vagueness

Returning to the weakly paracomplete logic of supervaluationist theory, as we saw, it contains a non-truthfunctionality account of disjunction

---

[26] Comp. Sainsbury (1989), Burgess (1990), Tye (2000), Morreau (2002), Rosen and Smith (2004), Hyde (2008). Similarly, the indeterminacy can be compositional, temporal or modal.

[27] Comp. Sider (2003), an argument based on a representational account of vagueness.

[28] "Nothing has any sort of shady presence", 237.

and a non-standard meaning of "there is". If we are not willing to endorse their consequences, then how must be a logic of vagueness, what requirements must it satisfy?

First of all, the standard sorites paradox is to be qualified as unsound, for its conclusion is false, therefore one of its premises is *not true.* Evidently, if such a premise would be declared false, then a problem immediately arises, for "$\varphi(a_1)$" is true (by supposition) and no conditional premise "$\varphi(a_k)\rightarrow\varphi(a_{k+1})$" can be false, otherwise "$\varphi(a_k)$" will be true and "$\varphi(a_{k+1})$" false, both *determinately* so, hence "φ" would be not soritical, having a sharp boundary. Therefore, a logic of vagueness cannot declare one of the premises false. It will be *not true.* And because not all the premises can be true, it results that such a premise is *indeterminate*. In moving from a_1 to a_n we find the cases of indeterminacy in borderline points of φ, a case in which the conditional has a true antecedent and an indeterminate consequent, i.e. it will be an indeterminate conditional; and, again, in passing from borderline points to determinate not-φ, the conditional will be indeterminate. Between them the conditionals will be true, having both the antecedents and the consequents indeterminate. This is the case of *first-order vagueness*.

Secondly, in passing form a_1 to a_n we do not discover a limit separating determinate points from the borderline points. That is, such limits are not determinately detectable and, therefore, the *higher-order vagueness* must be admitted.

Thirdly, a uniform treatment of different forms of sorties is required, by adopting a truth-functional account of logical meanings.

Is there such a logic?

Many proposals were made, for solving vagueness paradoxes and some semantic paradoxes. As a logic of vagueness, an important proposal is Hyde´s strongly paracomplete, truth-functional logic L_3.[29]

Shortly, this logic is a three-valued one, in which the intermediate value, ½, is a technical device, "a convenient fiction", expressing the idea of

[29] See Hyde (2008), Ch 7.

a gap: "neither true nor false", and represented as a third value.[30] If v is a valuation function mapping sentences on to the truth-set {1, ½, 0}, then $v(\neg \alpha) = 1 - v(\alpha)$, $v(\alpha \wedge \beta) = \min\{v(\alpha), v(\beta)\}$, $v(\alpha \vee \beta) = \max(v(\alpha), v(\beta))$.

$v(\forall x \varphi(x)) = \min\{v\varphi(x), x \in \text{Dom}\}$, $v(\exists x \varphi(x)) = \max\{v\varphi(x), x \in \text{Dom}\}$, $v(D\alpha) = 1$ if $v(\alpha)=1$; 0 otherwise, where "D" is the operator "determinately", $v(I\alpha)=1$, if $v(\alpha)=½$ where "I" is the indeterminacy operator; $v(\alpha \rightarrow \beta)= 1$ if $v(\alpha) \leq v(\beta)$, and $1-(v(\alpha)-v(\beta))$, otherwise, where "→" is the Lukasiewicz conditional. In this logic *modus ponens* is a valid rule of inference, and disjunctive syllogism, adjunction and *ex falso quodlibet* are also valid. By contrast, the law of excluded middle and non-contradiction fail. Similarly, the conditional proof, the deduction theorem, the contraposition and r*eductio ad absurdum* do not hold.

This three-value (gap) approach does hold also for higher-order vagueness. Even if a sentence is indeterminately indeterminate, the values "true", "false" and "neither" are sufficient for treating such cases. The vagueness of "vague" or of some other predicates does not require the introduction of additional values, other than the three mentioned above. According to Hyde's view, (2008), 198, higher-order vagueness in the object-language can be treated by using a *vague* higher-order metalanguage.

> Just as the object-language is vague, so too the metalanguage. Higher-order vagueness does not require the postulation of infinitely many truth-values but instead can be accommodated by recognizing that vagueness is a feature not only of the object-language, but of the infinite hierarchy of metalanguages as well. In this sense, higher-order vagueness of a language can be accommodated by vagueness in higher-order languages**.**

[30] "The third 'value', the gap reflects a distinct semantic category into which sentences may fall. It captures the distinct truth-value status of a vague sentence but is not itself a truth-value" (2008), 200.

Vagueness and related matters

As we saw above, the vagueness of a predicate means "there is no determinate fact of the mater" as to whether the predicate applies or not. That is, the vagueness is a form of indeterminacy in application of the predicate. Treated with the notions of classical logic if gives rise to paradoxes of vagueness (sorites). As is well now, there are some other predicates (e.g. "instantiates", "true") that generate paradoxes and, therefore, some kind of indeterminacy in their applications. Such a connection between both kind of paradoxes though not generally accepted, makes the issue of important investigations.[31] We do not pursue to analyse these matters here, but only to point out some logical aspects of this connexion, regarding the indeterminacy as a common feature and the paracomplete logic as a common treatment of it.

The indeterminacy arises in paradoxes of self-reference. We give some example.

The König paradox

If L is a given language and M is a set, then M is definable in L if there is a formula of L, $\alpha(x)$, with only x free, such that for every n natural number the following holds

$$n \in M \text{ iff } \alpha(n) \text{ is true}$$

Or, similarly, and more generally, an entity e is definable in L if $\alpha(x)$ *is true*[32] of e and only of e. If L contains only a finite list of symbols, then in L only a countable infinity of expressions can be constructed and a finite number of expression of length less than 100.

The König paradox involves the following ideas: a) there are uncountably many ordinal numbers, and b) the ordinal numbers fall into a natural well-ordering. Now, the construction of the paradox runs thus: since there are only countably many expressions of L, there are ordinal numbers that are not definable in L (by a)). According to *the least ordinal principle*

---

[31] E.g. McGee (1989), (1991), Tappenden (1993), Field (2008).

[32] We suppose that L contains "true" or "true of", in that case "definable in L" is itself definable in L.

(motivated by b)), there is a unique smallest ordinal not definable in *L*, let it be *o*. We have therefore.

(*) *o* is the smallest ordinal undefinable in *L*

(**) "*x* is the smallest ordinal undefinable in *L*" is true only of *o.*

To (**) corresponds a formula of *L*. Then the paradox is: *o* is the smallest ordinal undefinable in *L*, but *o* is definable in *L* by "*x* is the smallest ordinal undefinable in *L*".

In the same way can be constructed *Berry's paradox*. It is based on the notion of *least number principle*, and uses the idea of definability by an expression of *L* with length less than 100. Let the following expression be:

"the least natural number not definable by an expression whose length is less than 100".

The paradox results from the fact that such a natural number has just been defined by using an expression with fewer symbols that 100.

Grelling's paradox

A predicate is called *heterological* if it is not true of itself i.e. it has not the property that is expresses. The predicate "long" is not long, that is, it is heterological. The predicate "abstract" is itself abstract, hence it is not heterological. The paradox arrives when we ask if "heterological" is heterological. The paradoxicality can be derived in the following way:

" Heterological" is true of " heterological" iff " heterological" is not true of itself.

equivalently,

"Heterological" is true of itself iff "heterological" is not true of itself, an expression of the form "α iff $\neg \alpha$" being thus obtained.

Russell's paradox (for sets)

It is a construction similar to Grelling's, and results by considering the following expression:

Let *M* be the set of all sets that are not members of themselves. Is *M* a member of itself or not?

It is easy to see that each answer to this question entails the opposite.

Russell's paradox (for properties)

This paradox is similar to the previous, using in this case the property of "not instantiating *itself*" instead of expression "is not a member of itself", obtaining

The property of not instantiating itself instantiates itself iff it dosen't instantiate itself, a phrase of the form "α iff $\neg\, \alpha$", contradictory in classical logic.[33]

How can be treated such paradoxical constructions?

According to classical logic, regarding Kőnig paradox, the problem arrives in passing from (*) to (**). If this step is blocked, then the paradox disappears. Similarly, for Berry's paradox. But another diagnosis seems to be more plausible: the real problem is with (*) based on the unrestricted form of least ordinal principle, or of least number principle (in Berry's paradox). These principles do not follow from the idea of a well-ordering. What show both paradoxes is that the notion of definability has not sharp boundaries. Therefore, if definability manifests a kind of vagueness and (*) is based on the least ordinal/number principle, then the problem is with these principles. Why?

We take as exemple *the least number principle* (or well-orderedness of the natural numbers). According to this principle, if there exists a natural number x such that $\alpha(x)$, then there exists *a least* such x, call it y. Symbolically,

If $\exists x \alpha(x)$, then $\alpha(y) \wedge \forall z\ (z < y \supset \neg\, \alpha(z)$

Considering that a man with 1 hair is bald and that such a man with 10^4 hairs is not bald, by least number principle, there is a number y such that he was bald with y hairs and not bald with z hairs, for any z < y. As we saw by analysis of sorites this conclusion seems to be unacceptable. Therefore, these ingredients: the limits (1 hair, 10^4 hairs) plus least number principle (supposing an ordering 1, 2, …, 10^4) lead to a counterintuitive idea of the

[33] The solution to Russell's paradox for sets is well-known: there is no set whose members are the sets which are not members of themselves, a solution motivated by the hierarchial image of sets. This shows us an asymmetry between this paradox and the paradox for properties. According to Gödel, "[t]here never were set-theoretic paradoxes, but the *property-theoretic* paradoxes are still unresolved" (in J. Myhill (1984), 129-143).

existence of a sharp boundary between bald and not bald. To avoid such a conclusion a restriction in the application of *tertium non datur* to vague notions is needed, in the sense that we cannot assume that with *y* hairs he was either bald or not bald. In this case the least number principle will be weakened to the following form:[34]

If $\exists x\alpha(x) \wedge \forall z\,(z < y \supset (\alpha(z) \vee \neg\alpha(z))$, then $\alpha(y) \wedge \forall z\,(z < y \supset \neg\alpha(x))$

Therefore, it is inappropriate to say that with *y* hairs that person was either bald or not bald. And it is inappropriate to say that there exists the first *y* such that he becomes bald.[35] "It is inappropriate to say"[36] does not mean "there exists" or "there is not", for the negation of a fuzzy sentence is also a fuzzy sentence. And this locution does not imply an epistemic reading, in the sense that even if it is inappropriate to say something about such an *y*, however such an *y* there exists. For in such a case the applicability of *tertium non datur* will be reinstalled and a sharp boundary between bald and not bald will be fixed.

Therefore, the vagueness of some notions requires the weakening of the least number principle. The same consideration holds for the least ordinal principle, used is Kőnig paradox. Both principles will be reduced to their classical form if the formula α satisfies *tertium non datur.* Hence, the solution adopted to these paradoxes is the use of a paracomplete logic, in which *tertium non datur* has no universal applicability.

Regarding Russell's paradox for properties a paracomplete solution implies non applicability of *tertium non datur* to "circular" predicates like "instantiates". That is, the disjunction *Prop either instantiates itself or doesn't* is not accepted (where *Prop* is Russell's property). Finally, the rejection of *tertium non datur* can save Grelling's construction from paradoxicality. In both cases, Russell's and Grelling's, what is obtained is

---

[34] Field´s proposal, in (2008), 101.
[35] "The range where this is fuzzy will itself be fuzzy"; Field (2008), 101.
[36] In Field´s account this locution has not an objective meaning, but rather "[…] when I speak of a claim as inappropriate all I really mean to be doing is rejecting the claim" (2008), 101.

an equivalence of the following form: $\alpha \equiv \neg \alpha$. It is contradictory in classical logic, for assuming the validity of *tertium non datur* a sentence of the form $\alpha \wedge \neg \alpha$ can be derived:

1. From $\alpha \equiv \neg \alpha$ and α, $\alpha \wedge \neg \alpha$ can be obtained
2. From $\alpha \equiv \neg \alpha$ and $\neg \alpha$, $\alpha \wedge \neg \alpha$ can be derived

According to the Rule of Reasoning by Cases , if Γ, α_1 imply γ and Γ, α_2 imply γ , then from Γ, $\alpha_1 \vee \alpha_2$ the formula γ is obtained. Therefore, from 1 and 2 , by this rule, from $\alpha \equiv \neg \alpha$ and $\alpha \vee \neg \alpha$, the formula $\alpha \wedge \neg \alpha$ results. With the assumption $\alpha \vee \neg \alpha$, from $\alpha \equiv \neg \alpha$ results $\alpha \wedge \neg \alpha$. Without *tertium non datur* (in the de Morgan logics, in which $\alpha \equiv \neg \neg \alpha$ holds, for example) $\alpha \equiv \neg \alpha$ is not contradictory.[37]

In the same way, by rejecting the validity of *tertium*, other paradoxes can be blocked, Curry's paradox [38] , for example.

By Diagonalization Lemma a sentence α can be constructed, equivalent to the sentence.

True (α) →The earth is flat, or

True (α) → $\perp$, where "$\perp$" is "The earth is flat"

The following derivation holds:

1. $\alpha \equiv$(True (α) $\supset \perp$)

2. True (α) ≡ (True (α) $\supset \perp$; 1, assuming the intersubstituivity of α with True (α)

3. True (α) $\supset$ (True (α) $\supset \perp$), 2, by classical logic

4. (True (α) $\wedge$ True (α)) $\supset \perp$; 3 by importation rule

5. True (α) $\supset \perp$; 4

6. (True (α) $\supset \perp$) $\supset$ True (α); 2, by classical logic

7. True (α); 5, 6, *modus ponens*

8. $\perp$; 5,7, *modus ponens*.

---

[37] In classical logic $\alpha \equiv \neg \neg \alpha$ holds, hence $\neg(\alpha \equiv \neg \alpha)$ holds, for an inference from $\alpha \equiv \neg \beta$ to $\neg(\alpha \equiv \beta)$ is valid. But $\neg(\alpha \equiv \neg \alpha)$ does generate a contradiction with the conclusion of mentioned paradoxes. Thus what is sought is a paracomplete logic in which this inference is not valid.

[38] Curry (1942).

The paradox can be blocked by cutting out the step from 3 to 4; that is by restricting the importation rule. And this is the case in the Lukasiewicz[39] logics, shortly described in a previous section. The semantics of such a logic is a Kleene semantics plus the specific semantics for the conditional. It is easy to see that the intersubstituitivity of α with True (α) is preserved, for α and True (α) have the same value: True (α) $\rightarrow \perp$ has the value of 1 – True (α), for the value of $\perp$ is 0. Therefore, for α=½, True (α), α and 1-True (α) have the same value. But for these values the formula in 3 has the value 1 and the formula in 4 has the value ½. Therefore, the reasoning 1-8 is not valid, for the importation rule does not hold.[40]

Conclusion

Undoubtedly the vagueness raises a wide variety of questions, logical and philosophical. The sorites shows that this phenomenon cannot be treated only with the means of classical logic, and that even the finding of the adequate logic of vagueness depends on some philosophical questions, regarding the sources of vagueness. If vagueness is merely representational, then by a scientific precisification of our language it simply disappears, without loss of the descriptive power of the language (Russell's view). But sometimes the elimination of vague terms of a language, these being inconsistent, can have drastic ontological consequences: the corresponding objects do not exist (Unger and Wheeler). Or, in a strong realist view (epistemicism), though a term like "bald" is vague, the limit between bald and not bald there exists, the vagueness having no ontological roots. Or, by admitting that vagueness is, in many cases ontological, a number of questions arise regarding the idea of identity. This knitting logic-ontological is a note of any treatment of the phenomenon of vagueness.

The same strong connection can be pointed out in respect of the status of higher-order vagueness, of the soundness and relevance of Evans

---

[39] Lukasiewicz and Tarski (1930).

[40] Unfortunately this diagnosis is not "universal", in the sense that, for other Curry's sentences the number of required logical values increases. At the limit an infinite valued semantics will be used.

argument, or of the notion on vague existence. Similarly, for the question regarding the proper logic of vagueness. It seems that it must be a paracomplete one, but which of them? And, finally, has the phenomenon of vagueness an extension such that it covers the situations generated by paradoxes?

For all of these questions there are arguments *pro* and *contra*. And it is similarly true that in many cases is not easy to give an adequate answer.

References

Arruda, A. (1989), "Aspects ot the Historical Development of Paraconsistent Logic", in G. Priest, R. Routley and J. Norman (eds.), *Paraconsistent Logic: essays on the inconsistent*, Philosophia Verlag, 99-130.

Burgess, J.A. (1990), "Vague Objects and Indefinite Identity", *Philosophical Studies*, 59, 263-287.

Cantini, A. (1990), "A Theory of Formal Truth Arithmetically Equivalent to ID_1", *Journal of Symbolic Logic*, 55, 244-259.

Carnap, R. (1950), *Logical Foundations of Probability*, Chicago UP

___. (1966), *Philosophical Foundations of Physics*, ed. M. Gardner, Basic Books.

Church, A. (1960), "Vague", in D. Runes (ed), *Dictionary of Philosophy*, Philosophical Library, New York, 329.

Curry, H. (1942), "The Inconsistency of Certain Formal Logics", *Journal of Symbolic Logic* 7, 115-117.

Dummett, M. (1975), "Wang´s Paradox, *Synthese*, 30, 301-324.

Evans, G. (1978), "Can There Be Vague Objects?", *Analysis*, 38, 308.

Field, H. (2008), *Saving Truth from Paradox*, Oxford UP.

Fine, K. (1975), "Vagueness, Truth and Logic", *Synthese*, 30, 265-300.

Friedman, H; M. Sheard (1987), "An Axiomatic Approach to Self-Referential Truth", *Annals of Pure and Applied Logic*, 33, 1-21.

Garrett, B.J. (1988), "Vagueness and Identity", *Analysis*, 48, 130-134.

Geach. P.T. (1980), *Reference and Generality*, 3rd ed , Cornell UP.

Hyde, D. (1994), "Why Higher- Order Vagueness is a Pseudo-Problem, *Mind*, 103, 35-41.

___. (2003), "Higher-Orders of Vagueness Reinstated", *Mind*, 112, 301-305.

___. (2008), *Vagueness, Logic and Ontology,* Ashgate.

Keefe, R. (1995), "Contingent Identity and Vague Identity", *Analysis*, 55, 183-190.

___. (2000), *Theories of Vagueness*, Cambridge UP.

Lewis, D (1988), "Vague Identity: Evans misunderstood", *Analysis*, 48, 128-130.

Lukasiewicz, J., A. Tarski (1930), "Investigations into the Sentential Calculus", in A. Tarski (1956), *Logic, Semantics, Metamathematics*, Oxford.

McGee, Vann (1989), "Applying Kripke´s Theory of Truth", *Journal of Philosophy,* 86, 530-539

___. (1991), *Truth, Vagueness, and Paradox*, Hackett.

Morreau, M. (2002), What Vague Objects are really like", *Journal of Philosophy*, 99, 333-361.

Myhill, J. (1984), "Paradoxes", *Synthese,* 60, 129-143.

Parsons, T. (2000), *Indeterminate Identity*, Oxford UP.

Quine, W.V.O. (1960), *Word and Object*, MIT Press

___. (1981), "What Price Bivalence", *Journal of Philosophy*, 78, 90-95.

Rossen, G.: N.J.J. Smith (2004), "Worldly Indeterminacy: a rough guide", *Australasian Journal of Philosophy*, 82, 185-198.

Russell, B. (1923), "Vagueness", *Australasian Journal of Philosophy*, 1. 84-92.

Sainsbury, R.M. (1988), *Paradoxes*, Cambridges UP.

___. (1989),"What is a Vague Object?", *Analysis*, 49, 99-103.

___. (1995), "Why the World Cannot be Vague", *Southern Journal of Philosophy*, 33, 63-81.

Sider, T. (2003), "Against Vague Existence", *Philosophical Studies*, 114, 135-146.

Sorensen, R. (1985), "An Argument for the Vagueness of 'Vague'", *Analysis*, 45, 134-137.

___. (1998), "Ambiguity, Discretion and the Sorites", *Monist* 81, 217-235.

___. (2001), *Vagueness and Contradiction*, Oxford UP.

Tappenden, J. (1993), "The Liar and Sorites Paradoxes: Toward a Unified Treatment", *Journal of Philosophy*, 90, 551-577.

Tye, M. (1990), "Vague Objects", *Mind,* 99, 535-557.

___. (1994), "Sorites Paradoxes and the Semantics of Vagueness", in J. Tomberlin (ed), *Philosophical Perspectives* 8, 189-206. (2000), "Vagueness and Reality", *Philosophical Topics,* 28, 195-209.

Unger, P. (1979) a, "I do not Exists", in G.F. Macdonald (ed), *Perception and Identity*, Macmillan, 235-251;

___. b, “There are no Ordinary Things”, *Synthese*, 41, 117-154;

___. c, “Why there are no people”, *Midwest Studies in Philosophy*, 4, 177-222.

___. d. (1980), “Scepticism and Nihilism”, *Nous*, 14, 517-545

Varzi, A. (2003), “Higher-Order Vagueness and the Vagueness of ‘Vague’”, *Mind*, 112, 295-298.

Wheeler, S.C. (1979), “On that Which is Not”, *Synthese,* 41,155-173.

Wiggins, D. (1986), “On Singling out an Object Determinately” in P. Pettit and J. McDowell (eds), *Subject, Thaught and Context*, Oxford UP, Ch 6.

Williamson, T. (1994), *Vagueness,* Routledge.

Wright, C. (1987), “Further Reflections on the Sorites Paradox”, *Philosophical Topics*, 15, 227-290.

___. (1992), “Is Higher-Order Vagueness Coherent?”, *Analysis*, 52, 129-139.

Irrealistic Pluralism, Extensionalism, and Existence

Mark S. McLeod-HARRISON [*]
George Fox University

Abstract:

This paper argues that any pluralism rooted in noetic irrealism must solve two problems—the "anything goes" challenge and the "consistency" challenge. In order to solve those problems, however, it is argued that no pluralist of this type can be an extensionalist but rather must hold that existence is a (real) property.

Keywords: irrealism, antirealism, extensionalism, intensionalism, existence, pluralism, Goodman, Lynch.

Most ontological pluralisms are rooted in some sort of noeticism where the mind contributes significantly to the way the world is. By a significant contribution of the mind to the world I mean that human noetic work makes or shapes the world by contributing something more than artifacts and ideas. We humans somehow (and the story differs with the theory) make or shape the ontology of the world not simply by making artifacts and ideas but by making or shaping the natural and perhaps the supernatural as well. We make or shape not only cars but stars. Call this "noetic irrealism." While noetic irrealism itself does not entail ontological pluralism, it is arguably the best way to generate pluralism or, short of that, to explain pluralism.

Pluralists use different terminology to pick out the antithetic ways the world is. My terminology is as follows: I will say there is a singular World (using the upper case) made and/or shaped by human noetic

[*] E-mail: mmcleodharriso@georgefox.edu.

structures into various ways the World is, using the term "world" (lower case) to pick out those antithetical ways. A realist view of the World is the position that there is a singular World that is not, except in some obvious ways (such as the thoughts contained therein), made or shaped by human noetic work. When I refer to the realist World I will use quotation marks, as in "the World." Furthermore, for convenience I will use the term "pluralism" as shorthand for "noetically irrealist pluralism." Two intertwined challenges face pluralism. My thesis is that although these challenges must be met, a pluralism committed to extensionalism cannot meet them, nor can a pluralism denying that existence is a (real) property.

The Challenges

One issue facing pluralism is how to avoid what we might call "the anything goes challenge," roughly the claim that once the door to pluralism is open, we can create worlds any way we wish or with no limits. Put another way, what is to keep a more-or-less modest pluralism or relativism with supposed or assumed limits from falling all the way into a radical relativism with no limits, what we might refer to as an extreme antirealism with total subjectivity? Some pluralists make explicit that there are limits on world-creation. Nelson Goodman, for example, describes his position as a "radical relativism under severe restraints."[1] Michael Lynch also claims that there are limits to how the worlds can be or as he puts it, "one can be a pluralist without having to believe that anything goes."[2] But whence these limits?

A second difficulty with pluralism is the consistency challenge. Suppose there are two conceptual schemes (or perspectives or versions, etc.) eventuating in two worlds, W_1 and W_2. Suppose further that A is true in W_1 and –A true W_2. Thus A and –A are equally true on the pluralist's grounds. Now either the descriptions of W_1 and W_2 are consistent with each other or not. If not, then it seems the law of noncontradiction is violated and then just anything will go. To avoid that end, the pluralist will say that A and –A

[1] See Nelson Goodman *Ways of Worldmaking*, Indianapolis: Hackett Publishing, 1978, x.
[2] Michael Lynch, *Truth in Context*, Boston: MIT Press, 1998, 3.

are consistent with one another and that will be accomplished by relativising truths to worlds growing out of conceptual schemes: A is relative to W_1 and –A to W_2. This consistency, however, turns out to be problematic. If A and -A are consistent, then they must be expressing the same absolute truth in different languages or they are concerned with different subject matters. But the absolutist can accept that. One should then wonder, however, what motivates the pluralist position, since pluralism doesn't have much purchase on the World that can't be easily handled by the absolutist.

These two challenges are linked. One of the most important limits on how things are (or how a world can be built) is the law of noncontradiction. Denying the law of noncontradiction opens the worlds in the wildest of ways. Indeed, it is arguable that it is the law of noncontradiction itself that breaks the World into worlds and hence generates conflicting world-descriptions.[3] The limits on what will go in a given world are linked to how the pluralist is to keep the various incompatible worlds separate from one another so as to allow the worlds to be built without direct contradiction across world descriptions. Of course, all this is done while maintaining that the world descriptions are, in the end, incompatible with one another. In short, the anything goes challenge, at its worst, admits that the law of noncontradiction doesn't hold. To avoid this eventuality, the pluralist attempts to relativise truth to worlds. But that raises the consistency challenge.

The pluralist needs solutions to both these challenges. I think solutions are available. The question is how to meet the challenges. I believe the anything goes challenge cannot be met with an extensionalist framework without falling into subjectivist caprice. Furthermore, the consistency challenge, I believe, can be met only within an intensionalist framework. Thus intensionalism provides the basis for a solution to both challenges.

---

[3] See my *Make/Believing the World(s): Toward a Christian Ontological Pluralism*, forthcoming McGill-Queens University Press, 2009.

Extensionalism

One way to place a limit on how worlds can be built is to hold steadfastly to extensionalism. Extensionalism is the view presupposing that the use of terms is completely determined by what falls under the terms in the (or a) actual world. An extensional meaning or sense is given by listing or otherwise indicating the (actual) things that are referred to by the term. Since "Morning Star," "Evening Star" and "Venus" all refer to the same thing, the extensional meaning of those terms is identical to the thing referred to by each, which in this case, is one particular extraterrestrial object.

Some pluralists are extentionalist. Goodman, for example, roots his nominalism in extensionalism, assuming that the only things that exist are individual things.[4] Some pluralists might, on the other hand, be extensionalist platonists. On such an account, the properties and kinds countenanced by the platonism can admit no more than what actually is. The difference between the nominalist and the platonist here is simply that the nominalist recognizes only individuals while the platonist also recognizes kinds and properties. For an extensionalist platonist, however, kinds and properties exist (only) in the actual world(s). There are no intensional objects.

W. V. O. Quine's physicalism is a kind of extensionalist platonism, for it tells us that the only kind of individual that exists is a physical one. He must, therefore, have some means of admitting only physical kinds and not others. To do so, he must admit some sort of property, viz., the property of being a physical object. But for Quine, as for any extensionalist, it does not follow that there are properties independent of the way the World is or worlds are. There is only a property of being a football in virtue of footballs being actual. Merely possible footballs won't do it.

The nominalist, in contrast, remains open to all sorts of things counting as individual objects, including the physical, the phenomenological, and so on. For example, Goodman writes:

---

[4] See Goodman, *passim*.

> I am sometimes asked how my relativism can be reconciled with my nominalism. The answer is easy. Although a nominalistic system speaks only of individuals, banning all talk of classes, it may take anything whatever as an individual; that is the nominalist prohibition is against the profligate propagation of entities out of any chosen basis of individuals, but leaves the choice of that basis quite free. Nominalism of itself thus authorizes an abundance of alternative versions [or worlds] based on physical particles or phenomenal elements or ordinary things or whatever else one is willing to take as individuals. Nothing here prevents any given nominalist from preferring on other grounds some among the systems thus recognized as legitimate. In contrast, the typical physicalism, for example, while prodigal in the platonistic instruments it supplies for endless generation of entities, admits only one correct (even if yet unidentified) basis[5].

Whereas Quine allows for no properties but physical properties (taken in the extensionalist sense), Goodman allows for no properties at all, whether extensionally or intensionally understood. Quine's platonism thus comes via his commitment to individual *physical* objects rather than through his requiring a limit on how we construct things in terms of individuals in general. His physicalism commits him to a platonistic version of extensionalism. Goodman, in contrast, thinks nominalism is the proper limit on how we can construct the worlds we do and that limit is strictly in terms of individuals. No properties are allowed Goodman, although they are for Quine, but neither Quine nor Goodman countenance intensional properties.

Intensionalism

Intensionalism is the view that there are uses of terms beyond what is determined by what falls under the terms in the actual World or worlds. We can understand intensionalism to be a (more or less rich) platonism that countenances kinds and properties that may or may not have actual things falling under them. Thus, unicorns as possible objects (unicorns exist in some possible world) have the property of unicorness attached to them, and

---

[5] Ibid., 94, 95.

thus the class of unicorns, although empty in the actual world, still exists, as does the property "being a unicorn." For the extensional platonist, possible entities don't exist. If there are no unicorns, there will be no property of unicorness. A platonism saying that unicorness exists even in a world that has no unicorns is not extensional. Since horses exist for the extensional platonist, there is a property of horseness, but that property is given account simply by referring to the set of all the horses. There is nothing beyond the horses themselves—there is no horseness that floats free of the horses. If horses all ceased to be, so would horseness.

Extensionalism and Pluralism

For the pluralist who wants also to be an extensionalist, whether of the platonic or nominalistic sort, something like the following would hold. Any world created by human noetic work would be an actual world. If the pluralist were a nominalist, then any world would contain only individuals. If the pluralist were a platonist, then any world would contain only individuals, kinds and properties made up strictly of the entities found within the actual world created. So if there were individual trees, there would also be the property "being a tree." But there would be no property "being a tree" independent of the actual trees in that world. Call the philosopher committed to pluralism and to extensionalist nominalism, a pluralistic nominalist. Call the philosopher committed to pluralism and to extensionalist platonism, a pluralistic platonist.

What relationship would hold between the nominalism of the pluralistic nominalist and her noetic irrealism? The latter says that what exists depends in some way on the noetic feats of the human. There is no mind-independent reality. So, not only do kinds not rest on any ready-made (set of) properties but neither do individuals rest on any ready-made base. Individuals can be furniture-sized objects, physical particles, phenomenal experiences, ghosts, or minds. Kinds, if such there be, are completely reducible to individuals. They are no more than what Goodman calls "relevant" or "historical" kinds (more or less convenient ways of speaking). But how can the pluralistic nominalist talk either of kinds *or* individuals without talking about properties?

For the pluralistic nominalist, properties don't exist. Why? It can't be because of extensionalism, for extensionalism admits of some platonisms. It can't be that the worlds are "pre-made" containing only individuals, for that denies noetic irrealism. It appears to come down solely to the pluralistic nominalist's preference for making worlds with only individuals. The pluralistic nominalist is likely to say simply that a world built around nominalism is a well-formed world. There are, she might say, limits on what will go within a world. Nominalism is a limit with which she happens (by choice) to work.

But here it is not just nominalism that is a controlling feature but extensionalism as well. Enter the anything goes challenge. What is to stop the pluralistic nominalist from slipping into a radical antirealism where just anything goes? While extensionalism and nominalism would provide limits on how worlds can be built, why pick either? Why not an intensionalist view? Once the constraints of "the World" are removed, what is to keep the pluralistic nominalist from an extreme antirealism that eventuates in a near-total relativism? Extensionalism may appear to be one way to avoid an extreme radical relativism. However, I think it isn't up to the job because extensionalism is chosen no less than nominalism or platonism. A similar case can be made against the pluralistic platonist. What sets the limits on world making? Why these platonic properties (limited as they are to the actual worlds created within the extensionalism)? Won't just anything go here just as well?

Let's concentrate for now on the pluralistic nominalist. If such an irrealist is to allow for limits in world making, she needs something beyond a self-imposed extensionalism or a self-imposed nominalism. The reason is that an irrealist cannot make worlds containing only individuals without introducing kinds and properties, contrary to her commitment to exclude both extensional and intensional kinds. The pluralistic extensionalist simply cannot make a success of this claim.

For the pluralistic nominalist, as one moves from one world to another, one will have *only* different individuals. But if there are no properties, what makes any individual an individual? To answer this question, let me appeal to Michael Lynch's contrast between thin and thick

concepts. Roughly, a thin concept is one shared across at least some conceptual schemes whereas a thick concept appears in individual conceptual schemes as a more robust version of a thin concept. For example, philosophers of mind all share the same thin concept of mind, that is, they agree that they are speaking of the thing that thinks as they theorize about what is referred to by the term "mind." Yet the physicalist and the dualist may have very different thick concepts of mind. This mechanism allows for the distinct ontologies of pluralism. Using Lynch's terms, and returning now to the question "what makes an individual an individual?" we can say that the pluralistic nominalist has a robust or thick view of objects. That is, she takes the minimal notion of an object and fills it out in such a way that *objects can only be individuals.* Things that exist (objects) are never kinds or properties, no matter the world in which they appear. But here's the rub. In being a nominalist and hence filling out the notion of an object as only and always an individual (never a kind), the pluralist takes an ontological stand that challenges her irrealism.

Thick and Thin Individuals

Because the pluralistic nominalist is an extensionalist, we know there are no worlds containing merely possible objects. As such, each world is an actual world. Furthermore, the pluralistic nominalist, qua nominalist, is committed to each and every world containing only individuals. Her nominalism thus limits the ways things can be. But there is a problem here, for just as Quine's commitment to physicalism demands platonism, so the pluralistic nominalist's commitment to worlds of individuals requires platonism. Her commitment to worlds containing only individuals requires properties that hold across all the worlds.

The challenge can be stated as a dilemma. Either the concept of individual is minimal or it is robust. If it is minimal, it appears not to be distinct from the notion of object. Now for the pluralist, what counts as an object will vary world to world, conceptual scheme to conceptual scheme, ontology to ontology. Given noetic irrealism, objects can turn out (within a given world robustly conceived) to be individuals or kinds. But the pluralistic nominalist says that only individuals exist, *no matter what world.*

The upshot appears to be that just as there are objects in every world, so there are individuals in every world. In any world where there is anything at all, there are objects *qua* individuals.

If the concept of individual is not distinct from the concept of object, then the thinness of the concept of an individual (*qua* object) must remain open to there being *kinds* of individuals (*qua* objects), for what objects there actually are *given* irrealism turns out to depend on how one chooses to thicken the concept of an individual (*qua* object). The pluralistic nominalist can't be a *strict* nominalist—can't require it "ahead of time"—if she takes this branch of the dilemma. She must allow that objects can be kinds. So says her irrealism. But this undermines her commitment to nominalism for all worlds. In short, by standing fast with nominalism, the pluralistic nominalist affirms a platonism of individuals.

If she takes the other branch of the dilemma, viz., if the concept of individual is robust from the start, then it is already filled out by an explanation of what the properties of individuals are. A thick notion of the individual already brings with it some sort of platonism for we know that no matter what world we consider, it will have only what we might call "true individuals," that is, no kinds. Yet by the pluralistic nominalist's admission, what "properties" things have (how the properties are thickened) will vary world to world. Hence one would expect nominalism to hold in some worlds but not in others. However, it appears that she can't actually allow for strict nominalism, since then the concept of individuals already is a robust kind, a kind filled out with certain properties. The properties that demand individualism in fact entail at least one kind, viz., the kind "individual." And if we suppose, *per impossible*, that the concept of individual is minimal but turns out to hold only in one world, then the pluralistic nominalist must spell out how the statement of nominalism is true—i.e., why there are individuals—in one world but not in another.

A thin conception of individuals seems to be what is needed to get pluralism off the ground, for individuals show up in every world. Yet within any given world, the individuals always show up as objects attached to thick concepts, for if they don't, there is nothing to stop them, so to speak, from being something other than individuals (such as kinds or

properties). If nominalistic individuals are thick *qua* individuals and hence turn up in every world, then all worlds will contain properties and hence at least one kind, viz., the kind "individual." Individuals, understood as thick individuals, undermine nominalism, for "being an individual" is a robust property and there is a robust class of objects that turn up in every world. Individuals as thin individuals undermine strict nominalism on other grounds, for the very thinness demands an openness to what there is, including kinds.

In short, it appears that there are (robust) properties attached to the concept of individual and hence nominalism holds across worlds (that is, in the entire universe of worlds). But that is inconsistent with the irrealistic aspect of the pluralists which claims to be open to extensional platonism. I think in the end the pluralistic nominalist is not just open to platonism, she is forced into it. Just how strong is this platonism? Is it so strong as to entail intensionalism?

Extreme and Unrestrained Relativism

I believe the pluralistic extensionalist is restrained by little but her own choice and hence turns out to be much more radically relativistic than she will want to admit. The concepts the irrealist uses in making a world are not themselves fixed ahead of time by any way "the World" is. Rather they are created by the world-makers. There are, as such, no ready-made kinds, although there can be historical kinds developed out of habit over cultural history—something a quite a lot weaker than Quinean platonistic kinds. Hence the world-makers select from among various ways of organizing and classifying, etc. Once a world is made, the individuals that exist are actual. That is, the extension is set by the making.

Within that world, are there kinds of things? This appears to be answerable only by appealing to what is actually in the world (by pointing to its extensional content). Some pluralistic nominalists might suggest that all talk of kinds can be nominalized. There are no actual kinds, just individuals. In short, these pluralistic extensionalists will want to keep the classifying purely on the side of the noetic. Yet, one of the things we can do noetically is to move from nominalism to platonism. This is just a matter of

redrawing the circle to include kinds. But at what level does this occur, before or after the extension is set? The pluralistic extensionalist might say that choosing one way rather than another is consistent with her noetic irrealism. She might say she could allow talk of kinds (as a Quinean physicalist might) but on the terms she has laid out, don't these still have to be accounted for on purely extensional grounds? Likewise with a nominalist view of worlds. But what is it that makes the worlds actual? What provides the extensional grounds in the first place? All this happens on the noetic side of the world-making—as yet the world isn't made and there is no actuality. Whether one is a nominalist or a platonist is up to the world maker and not preset by "the World." Why should she pick nominalism or platonism then? There appears to be no good reason outside the world as made and thus outside the arbitrary choice of the world-maker.

In fact, why should the pluralist pick extensionalism over intensionalism? Isn't extensionalism created on the noetic side, just as nominalism is, or Quinean platonism? Here it seems the pluralist might demur, saying that intensionalism introduces terms that don't have actual referents but which nevertheless pick things out—possibilities, say, or properties of things that don't exist. Intensionalism brings with it a truck-load of strong platonism. In particular, intensionalism seems to introduce the notion of essential properties. But the notion of essential properties will undermine alternative ontological worlds by claiming that things are *by essence* one way and not another.

But perhaps intensionalism can be understood in a way that pluralism is allowed. In fact, perhaps allowing for intensionalism, or at least some version of it, can help the pluralist avoid an extreme relativism. If the pluralist wants to say that not just anything will go on irrealism, why not try intensionalism as a means of providing certain limits? In particular, why not allow for intensionalism and a rich account of possible worlds? There is a good reason, in fact, to move to intensionalism, for it is exactly what is needed to solve the consistency challenge.

Before I move to that, however, it is worth noting that if the pluralist continues to work under the limiting claim that there are no properties, there is, it seems to me, a straight-forward argument to the conclusion of extreme

and unrestrained relativism. Since there are no real properties, truth turns out not to be real property. Since a singular account of the concept of truth (at least on a minimal level) appears necessary to make sense of logical terms being constant, without a realist account of truth there is nothing across the various worlds to keep logical terms constant. One can thus conclude that there is no reason to assume even the limits of logic and hence no limit on world-making. But in addition to the apparent arbitrariness of choosing the basic framework of world-building, including the logical framework, I think it is a mistake to believe that a strict extensionalism can even allow for noetic irrealism and its concomitant pluralism. To see why, we can consider Lynch's response to the consistency dilemma.

Existence and Properties

Lynch's reply to the consistency challenge suggests that the propositions made true in W_1 and W_2 are relative to different conceptual schemes and therefore logically consistent. Yet it is clear that the pair of propositions A and -A are incompatible in the sense that if they were relative to the same scheme, they would be inconsistent. Indeed, Lynch claims that it is necessarily true that in every possible world where the propositions in question were relative to the same scheme, only one is true.[6] I believe Lynch's possible world solution to the consistency challenge is correct and, in fact, the most, if not the only, plausible way out of the challenge. It demands, I believe, a commitment to intensionalism. The framework of intensionalism, fortunately, opens to the door to a solution to the anything goes challenge as well, for intensionalism allows for logical terms and other important limits on world building.

I hold, along with Lynch and William Alston, that alethic realism is consistent with ontological pluralism. I also hold, however, that pluralism *requires* a realist account of truth. My belief derives from thinking about the consistency challenge. Unless the consistency problem can be resolved, pluralism can't get off the ground. But the best and perhaps only successful solution to the consistency problem is built on the notion of possible worlds.

---

[6] Lynch, 93.

This particular appeal to possible worlds implies a commitment to some sort of intensionalism, at least if we are to see a resolution to the consistency problem. Of course various nonintensional accounts of possibility have been proposed, but when it comes to the solution to the consistency problem, I believe only intensional accounts will do the trick. Furthermore, I think that intensionalism requires a realist account of truth. Without a realist account of truth wherein truth is a real property, what is it that makes statements about possible objects true?[7]

William Alston distinguishes between truth as a concept and truth as a property.[8] It's been noted that although we can easily see that our having the concept of gold does not necessarily entail that we understand the richer, more robust set of chemical properties of gold, it is not so obvious that we can have the concept of truth without having some details about the property of truth.[9] What is it to have the concept of truth unless we understand its most basic properties, for example, that "'p' is true iff p" or more accurately, "p"'s being true iff p? As Alston suggests, this formula itself indicates the minimal properties of truth. In having access to the minimal account or concept of truth, we have access to the fact that truth is a property, but as such we are not thereby committed to all the richness of more robust accounts of truth such as the correspondence theory.

But it's not only truth that is a property. In order to make a success of the Lynchian response to the consistency challenge, I believe existence needs to be a property as well. One reason often given for the notion that existence is not a property is that there are no nonexisting objects. Existence is not a property because if it were, then one could divide objects into two kinds, objects that exist and objects that do not. But objects that do not exist aren't objects at all. The idea seems to be that since objects that

---

[7] I want to note that Lynch provides a more or less extensionalist account of possibility when he tries to show pluralism consistent with metaphysical realism. See "Pluralism, Metaphysical Realism, and Ultimate Reality" especially pp. 71-78. The essay is found in William P. Alston, ed. *Realism and Antirealism* Ithaca: Cornell University Press, 2001.

[8] See William P. Alston *A Realist Conception of Truth*, Ithaca: Cornell University Press, 1996 for a fuller discussion of these matters.

[9] Lynch, 130,131.

don't exist aren't objects, "existence" is not a property. It's not as if there are two kinds of objects—existing ones and nonexisting ones. Objects simply are not members of a kind and, as such, there is no property that distinguishes them from members of other kinds.

Of course, not everyone agrees with this analysis. Consider Alvin Plantinga, for example:

> Among the properties essential to all objects is *existence*. Some philosophers have argued that existence is not a property; these arguments, however, even when they are coherent, seem to show at most that existence is a special kind of property. And indeed, it is special; like self-identity, existence is essential to each object and necessarily so[10].

It is, perhaps, no coincidence that self-identity and existence are found together in every object. Perhaps, indeed, self-identity is conceptually connected to existence and as such is one aspect of the essence of existence. But we need not take a stand on this issue here. But I am inclined to take Plantinga's view that existence is a special kind of property, viz., as an essential property of objects. So for Plantinga, all objects exist. With this, of course, the pluralist who rejects existence as a property agrees. But it doesn't follow from the fact that all objects exist that existence is not a property. Instead, as Plantinga notes, one could conclude that existence is a special sort of property. It is important to see, however, that this conclusion does not undermine pluralism. What if existence were a thin property? Existence could be a thin property and hence fluid enough to be filled out robustly in different worlds, which is what is needed for pluralism.

Perhaps the pluralist can be helped out too by the further suggestion that the concept of a property is fluid. There would be, then, both a minimal concept of property and (potentially) many robust accounts. There are two ways to go with this suggestion when considering the notion of the minimal concept of property. The first is this: Just as existence is not (it is supposed)

---

[10] Alvin Plantinga, "Actualism and Possible Worlds," in *The Possible and the Actual: Readings in the Metaphysics of Modality,* edited by Michael J. Loux, Ithaca, New York: Cornell University Press, 1979, 261.

a property, so are properties not properties. The second is that properties are properties. The first case sounds like an explicit contradiction while the second sounds merely tautological. This is not the case, however, when we rephrase the claims thusly: Properties considered minimally are not real (intensional) properties and properties considered minimally are real (intensional) properties. I'll argue in the remainder of this paper that the first view is problematic and therefore we are left with the second view.

Before we enter that discussion, however, an observation about the relationship of noetic realism/irrealism to intensionalism/extensionalism should be noted. Extensionalism, as we know, says that only actual things exist. Intensionalism, on the other hand, suggests that merely possible things exist as well. Whether properties are intensional or only extensional seems connected to the noetic realism/irrealism discussion in this way. If there are intensional objects that are merely possible objects then it seems that in some respects they must be noetically real objects. If not, then they depend upon some actual human (or other world-maker) for their being and so far forth the objects are not intensionally merely possible but rather the merely possible is built up out of some sort of actual concepts rooted in the actual thought of some actual human. The merely possible is no longer existent "on its own" as the intensionalists want but a sort of nonmodal actualism (such as we find in Nicholas Rescher).[11] But nonmodal actualistic accounts of possibility intend to do away with things such as intensional properties, that is, to make possible worlds talk consistent with extensionalism. Extensionalism seems to fit better with irrealism than does intensionalism. But we shouldn't hang too much on this last observation, for the fact that intensional objects, when merely possible, are noetically real does not entail that actual objects are noetically real.[12]

---

[11] See Nicholas Rescher, "The Ontology of the Possible," in Michael Loux, ed. *The Possible and the Actual* Ithaca: Cornell University Press, 1979.

[12] There is something counter-intuitive about the notion that intensional objects are noetically real while actual objects are irreal. I haven't the space here to explain how all this fits together, but I believe that intensional entities have their existence in God's mind (but not human minds) and so in some sense intensional objects are irreal as well, but divinely so rather than humanly so. See my *Make/Believing the World(s).*

Returning now to the discussion of properties, the first position noted above—that properties are not real (intensional) properties—could be seen to be motivated by the move that properties, like objects, are tied to what exists, so that just as there are no non-existing objects, so there are no non-existing properties. The minimal concept of a property might be "the features that makes objects what they are" but since, on our assumption, filling out what an object is depends on its properties, there will have to be robust concepts of properties as well. On the minimal understanding of properties, there is no property of being a property, just as there is, for example, no property of being an object or no property of existence. Nevertheless, we have a concept of property, a minimal one, (just as we do with object and existence) which will be filled out in a variety of ways in the many worlds we make. One robust account of properties is an intensional account—read "noetically realist account." A second account is more noetically nonrealist but therefore a more extensional account of properties.

But if we make this move, viz., the move to (minimal) properties not being (intensional) properties and hence leave open how to fill in the ontology of properties in various conceptual schemes, we must recognize that there are limits, if one wants to remain a pluralist. If one provides too strong a realist account on the robust level, it will turn out that there is only one way the World can be, viz., the singular way it is—"the World." This is parallel, one supposes, to filling out the minimal realism about truth as a strong correspondence theory. If one takes the concept of truth that way, it undermines any claims to a realist account of truth being compatible with various sorts of noetic irrealism. But there is a limit going the other direction too, for a radical noetic irrealism about properties undermines the pluralist's distinction between minimal and robust properties, for anything could then count as a property—anything will go and we are left with an extreme relativism. This is the charge I laid at the nominalistic pluralist's feet earlier. In the end, the notion of minimal concepts itself seems to rely on there being some property, no matter how thin, that makes a concept a concept. It's not clear that we can ever be free of properties, and noetically real ones at that. As such, not only must properties be real (intensional) properties but being an object and existing too must be properties, even if

minimal ones, should pluralism have any chance of avoiding extremely radical relativism.

We see this struggle in Lynch when he discusses Alston's distinction between truth as a concept and truth as a property.[13] How do we investigate the property of truth, via *a posteriori* or *a priori* methods? Unlike gold, which we can have a concept of without knowing its true makeup, it's less clear that we can have a concept of truth without knowing something about its properties. In fact, Alston and Lynch both say that robust accounts of truth will fill in the properties of truth but that any realist account of truth, even the minimal realist account, views truth as a property. This property must be a realist property—a property to which no human noetic/epistemic contribution is made, which both Lynch and Alston admit. But then doesn't the minimal concept of a property face the same issue? How is one to have the (minimal) concept of a property without recognizing that being a property is being a *noetically real* property? As such, minimal properties are real. And that implies that existence is a (real) property too. And as intensionally real, then noetically real as well.

What does that do to pluralism? Does pluralism need it to be true that existence is not a property in order to get pluralism off the ground? I don't think so, any more than it needs truth not to be property. One can be a minimal realist about truth, and a minimal realist about properties and existence (in the sense that each is a minimal property), and yet there be many actual ways the World is (that is, many worlds) that are filled out as the properties of objects are filled out beyond the minimal accounts. Recall that although merely possible intensional objects are noetically real, actual objects can be irreal. Here it is important to remind ourselves of the commitments to necessary truths and the usefulness of possible world talk that come along with Lynch's solution to the consistency challenge. Once we have admitted that even the minimalist account of truth requires real (intensional) properties then we are committed to existence being a property. Existence is a property, if a special one, such that all objects have the property of existence. Here, then, we have a commitment to

[13] Lynch, 131.

intensionalism. Intensionalism provides the grounds, in turn, for a rich modal actualism via which Lynch's solution to the consistency challenge can find success.

Conclusion

I've argued that pluralism is incompatible with a strict extensionalistic nominalism and further that pluralism requires, given the possible world solution to the consistency challenge, a realist notion of truth with existence being a real (intensional) property. I believe the general position holds for any noetically irrealist pluralism. Of course, significant segments of the argument assume that the possible world solution to the consistency challenge is the only viable one. However, I don't see an alternative available to the pluralist and thus I believe my suggestions hold.

Paradoxes of logical realism

Ionel NARIȚA*
West University of Timisoara

Abstract:

The thesis argued in this article is that logical realism generates paradoxes. Logical realism must be distinguished from other forms of realism such as ontological, linguistic or epistemic realism. Logical realism admits that the individual variables in a formula can be interpreted both by individual and predicative constants. In this way, logical realism disregards syntactic differences between the two types of constants. If, during the interpretation of variables, we take into account the syntactic constraints, and the logical realism is rejected, then, paradoxes such as *Impredicable*, or other types of paradoxes, are removed.

Keywords: Logical realism, Paradoxes, Logical syntax.

The realism had various forms during time, such as ontic realism, linguistic realism or epistemic realism. The ontic realism claims the thesis that mental entities, called *ideas*, have reality, sometimes more present that the material entities.[1] The ontic realism has several forms, some of them apparently opposed each another, as they reduce materiality to ideality or vice versa. The *idealism* or *spiritualism* sustains the preeminence of ideas, considering either that only the ideas exist or that they have a determinant role. On the contrary, the mecanicism, with its forms, *physicalism* or *physiologism*, reduces mental phenomena to the material ones. Despite their oppositions under certain aspects, the thesis that mental entities have reality, either spiritual or material, being realist doctrines, is commonly accepted.

* E-mail: inarita@litere.uvt.ro.

[1] Some authors call this kind of realism 'logical', because it admits the reality of the ideas. Lovejoy A. O., 2007, *The Revolt Against Dualism*, Read Books, New York, p. 114. In this study, the term 'logical realism' has another meaning in relation with the symbolic logic.

Ontic realism must not be a reductionism, but it can admit that mental entities are substantially different from the material ones, but they remain real. Such a doctrine is called *dualism*.

Against ontic realism raises *relativism*, which claims that mental entities are different from material entities, but the difference between them is not substantial, but temporal. Their own substance can't explain the spiritual phenomena, which do not have a different substantial support than the material phenomena. Spiritual phenomena take place in the present, while the material phenomena belong to the past, therefore, because the present is subjective, the spiritual phenomena cannot be objective, cannot be objects. Thus, it can explain why the mental phenomena are not accessible for an external observer.

The thesis of linguistic realism is that terms have a real correspondent in the same manner as names. For instance, terms like 'circle', 'man' or 'number 2' must correspond to objects, as to the names 'Aristotle' or 'Plato' correspond the objects named by them. This thesis has both syntactical and semantical consequences:

1) From the syntactic perspective, if there are objects corresponding to the terms, then the 'referent' of the term is the same for all users of the language, no matter what happens with it; therefor, the syntax of the sentences with terms as subject is the same with the syntax of the sentences which have names as subject. In fact, the linguistic grammar doesn't distinguish between those situations; from a linguistic point of view, the names (i.e. the proper names) are only a category of terms (nouns). In this way, the linguistic syntax sustains the realist thesis. The realist syntax supposes that expressions as 'Aristotle' or 'man' have the same behavior in sentences so that, according to syntactic criterion, we cannot distinguish between them. Both sentences 'Man is rational' and 'Aristotel is rational' would have the same subject-predicate structure.

2) From a semantical point of view, the linguistic realism generates the conclusion that, besides individuals, as names denotate, should exist 'ideal objects' or 'abstract objects' with a proper behavior. For example, while the truth-value of the sentences about individuals is variable in time, the truth-value of the sentences about abstract objects is the same in any

moment. While the truth-value of the sentence 'Aristotle is logician' has changed, being false during Aristotle's childhood, a sentence like 'The square has four right angles' remains true at any moment in time. From such a situation, the realists draw the conclusion that the abstract objects are unchangeable, and their properties are *essential*, therefore, they have true reality or necessary reality.

Nominalism is the opposite doctrine to the linguistic realism. The nominalism claims that only names have corresponding objects while terms have no real correspondent. The function of terms is to abbreviate the linguistic expressions. The abstract objects don't exist. Consequently, there are irreducible differences between names and terms. For instance, terms can't be subject in a sentence. Sentences must have names or notations as their subject; they can refer only to individuals or to other expressions.[2]

The *epistemic* realism sustains that theoretical terms have a real correspondent. The scientific theories contain, besides observational or empirical terms, which have a correspondent in our experience, some terms without correspondence in experience, called *theoretical* terms. Because the theoretical terms have no empirical support, the realist doctrine postulates that, beyond experience, there is a domain detached from our senses, containing objects corresponding to theoretical terms, called *reality*. In addition, the experience is only apparent and subjective while the reality has the property of objectivity. It follows that truth represents the correspondence with reality, not with experience.

Because the reality isn't accessible to our senses, it follows that only some people can reach the true knowledge by strange means. In this manner, the realism is equivalent with authority principle rejecting the tolerance principle. This is a strong reason to sustain realism for some people, because it allows to justify the claim to rule and lead in society.

Empiricism opposes to epistemic realism. Empiricism admits that theoretical terms have no correspondent both in experience and in reality;

---

[2] 'In nominalism, the basic thesis is that there are no universals and that there is only predication in language'. Cocchiarella N. B., 2007, *Formal Ontology and Conceptual Realism*, Springer, Dordrecht, p. 84.

there are no objects associated to the theoretical terms. The only function of theoretical terms is to abbreviate the linguistic expressions without supplementary knowledge. Some empiricists developed reductionist programs to substitute the theoretical terms with empirical terms conserving the empirical content of the theories. On contrary, the realists argued that theoretical terms are essential in a theory and they have the same cognitive importance as the empirical terms. The realists consider that the distinction between theoretical and empirical terms is arbitrary and artificial; both categories of terms play the same role in a theory; there is no syntactical distinction between them.

In conclusion, empiricism goes further than nominalism making a distinction between two categories of terms. The nominalist doctrine denies terms the role of the subject in sentences. Empiricism reaches the conclusion that theoretical and empirical terms satisfy in different ways their predicative function. While the empirical terms argue something with sense about the reference of the sentence, the theoretical terms have sense only if they can be reduced to the observational terms. According to empiricism, the expressions containing theoretical terms irreducible to the empirical terms are not, in fact, propositions, because they are not true or false. If a sentence contains theoretical terms without an empirical content, then it cannot be verified therefore, it doesn't have a truth-value; such a sentence doesn't transmit information and doesn't contain knowledge.

It follows that the main difference between realism and antirealism is that realism claims that expressions such as names, terms or even numbers, can play any syntactic role in a sentence, while the antirealism considers that the expressions must be distinguished from their syntax. An expression with certain syntax can take a place in a sentence determined by its syntactical structure. Any expression must be used only in conformity with its syntax.

The formulas of symbolic languages don't contain constant expressions like sentences, names or terms, but only logical constants and variable expressions. The expressions from symbolic languages are not interpreted by extralinguistical entities, as the expressions from natural language are, but they are interpreted by expressions from other languages,

for instance, by expressions from natural language. Consequently, the logical realism cannot refer to the correspondence between various kinds of expressions and reality.[3] For instance, symbolic languages don't contain terms about which we could ask ourselves if there is something objective corresponding to them or not; for a symbolic language reality has no importance. The expressions from symbolic languages are interpretable by other expressions so, logical realism raises the issue of the type of expressions that can be used for interpretation.

Starting from the fact that a variable supports different interpretations, logical realism admits that a variable could be interpreted by any expression from the natural or artificial language. For example, Gottlob Frege, one of the most consequent logical realists, doesn't fix any limitation for the interpretation. He considers the individual variables as empty places that can be filled by no matter what expressions belonging to a language, conserving the well formation. A formula as 'F(x)' must be read 'F( )' and it can receive various interpretations by filling the empty space between parentheses with a constant expression. For instance, the formula '$( )^2$+2=6'. The empty space can be interpreted by the expression '2', when it is obtained a true sentence: '2^2+2=6', but there is also possible to obtain a sentence using, as interpretation, the symbol of the sun: '$☼^2$+2 = 6' (even if it is a false sentence).

Bertrand Russell[4] showed that Frege's realism leads to paradox. If an empty place can be interpreted by no matter what expression, it can be substituted by the expression containing that empty space, too. For example, the empty space of the formula 'F( )' could also be interpreted by 'F( )', when we reach the expression 'F(F( ))', which is not saturate, so it isn't a sentence. If we would try to continue the interpretation, we would fallow in the *regresio ad infinitum* without filling the empty space. From this paradox,

---

[3] Brenner argues that Logic is not affected by the 'metaphysical' or 'linguistical' forms of realism because its independence from psychology. Brenner J. E., 2008, *Logic in Reality*, Springer, Dordrecht, p. 64.

[4] Though Russell brings arguments against Frege's realism, he remains, in many aspects, a logical realist. Cocchiarella shows that the Russell's realism is weaker in 1910 edition of *Principia* than in the previous edition, in 1903. Cocchiarella N. B., *op. cit.*, p. 86.

we can notice that the thesis of logical realism doesn't fit with the rule that the interpretation of the variables in a well formed expression must generate sentences.[5]

In their famous work, *Principia Mathematica*, Russell and Whitehead impose some restrictions, through the *theory of types*, to the interpretation of individual variables in order to avoid paradoxes like the precedent. Despite these precautions, the logical realism persists in *Principia Mathematica*; there are many situations when individual variables are interpreted by predicative or class constants.

A consequence of logical realism is the 'impredicable' paradox, built by Russell:[6]

1) We'll say about a predicate 'f(x)' that it is *predicable* if and only if it applies to itself, namely when 'f('f')' is taken into account. On the one hand, an *impredicable* predicate doesn't apply to itself, in other words, the predicate 'f(x)' is called impredicable if and only if '~f('f')' takes place. For instance, the predicate '*x* is red' is impredicable, because it is false in the situation '"red" is red', (a word cannot be colored). On the other hand, the predicate '*x* is a word' is predicable because the sentence '"word" is a word' is true.

2) In conformity with the previous considerations, we can define the predicate '*F* is impredicable' as it follows: $\text{imp}(F) =_{df} \sim F(\text{'F'})$, where 'F' is a predicative variable. In the same way the predicate '*F* is predicable' receives the definition: $\text{pred}(F) =_{df} F(\text{'F'})$. It is easy to notice that the relation: $\text{imp}(\text{'F'}) = \sim\text{pred}(\text{'F'})$ takes place.

3) Because the arguments of the predicate 'impredicable' are predicates, it appears the problem if 'impredicable' is predicable or not. Let's suppose it is predicable. In this case, it follows:

pred('imp') hyp. (1)
pred('imp') = imp('imp')
pred('imp') = ~pred('imp'), *contradiction*.

---

[5] Beaney M., ed., 1997, *The Frege Reader*, Wiley-Blackwell, London, p. 254.
[6] Cocchiarella N. B., *op. cit.*, p. 87.

If we now suppose, that ‘impredicable’ is impredicable, we reach again contradiction:

imp(‘imp’) hyp. (2)
imp(‘imp’) = ~imp(‘imp’), *contradiction*

We obtained the result that, despite the principles of logic, there are sentences both true and false, i.e., if the sentence ‘imp(‘imp’)’ is true, then it is false and reciprocally. Obviously, such a situation cannot be admitted in a logical system.

The source of the *impredicable* paradox consists in the admittance of the realist thesis that individual variables can be interpreted by any constants, including by predicative constants. The paradox can be avoided if we take into account the syntax of the constants that must correspond with the syntax of the interpreted variables. For instance, if ‘Fx’ is a monar predicative variable, the corresponding predicative constants can’t be f_1, f_2, …, f_n, but they must have a proper syntax: f_1x, f_2x etc., in other words, there is the possibility of confusion between predicative constants and individual constants.

The variables for predicate of predicates have the syntax ‘F(‘Fx’)’[7] so, the corresponding constants have the syntax ‘f(‘fx’)’. As logical operators and connectors are interpreted according to their syntax, it must be admitted that the variables have to be interpreted only if their syntax allows such an interpretation. For instance, the expression ‘p&vq’ isn’t a formula of propositional logic because the connectors like conjunction and disjunction are not used correctly. In the same manner, it must reject all cases of substitution or interpretation when the syntax constraints are violated.

[7] This discussion about monar predicate can be extended to other types of predicates with a similar result.

If we take into account the proper syntax of the constants, the *impredicable* paradox disappears. The predicate 'impredicable' is a predicate of predicates so, its correct definition is:

$$imp(\text{'Fx'}) =_{df} \sim F(\text{'Fx'}) \quad (3)$$

and the definition for *predicable* takes the form

$$pred(\text{'Fx'}) =_{df} F(\text{'Fx'}) \quad (4)$$

If we are asking ourselves if 'impredicable' is predicable or not, in order to answer such a question, we must interpret the predicative variable 'Fx' by predicative constant 'imp' in conformity with its syntax. The following relation is obtained:

$$imp(\text{'imp('Fx')'}) = \sim imp(\text{'imp('Fx')'}) \quad (5)$$

We notice the persistence of variable 'Fx'; we didn't obtain yet a sentence about which we could ask if it is true or not, because the expression (5) contains free variables. If we try to interpret 'Fx' again in the expression (5), we obtain *regresio ad infinitum*:

$$imp(\text{'imp('imp('Fx')')'}) = \sim imp(\text{'imp('imp('Fx')')'}) \text{ etc.} \quad (6)$$

If we don't fall in the trap of the realism and we take into account the expressions from logical syntax, the *impredicable* paradox disappears, proving that the predicate 'imp('Fx')' doesn't apply to itself, because 'imp' is a 'predicate of predicates' while its arguments must be simple monar predicates. A predicate of predicates applies to the predicates, no to the predicates of predicates. In the same manner, a predicate applies to individual constants and it is inapplicable to other predicates etc.

The neglect of the logical syntax, typical for realism, is the source of other paradoxes or paradoxical situations. For instance, we may ask about the reference of a physical law, as the law of uniform motion, $s = (v \times (t-$

$t_0))+s_0$. If we take into consideration the appearing constants in this formula, then the variables must be interpreted by numbers; if not the adding and multiplying have no sense. It follows that this law must be read: 'Number *s* is equal to the product between the number *v* and the number $t-t_0$, added to the number s_0'. Someone could understand that physical laws are referring to numbers. In fact, the law of uniform motion refers to physical bodies, but the previous formula doesn't say that. The correct formulation of this law, according to the logical syntax, is:

$$(s(t_0)Sx \ \& \ vVx) \supset ((v\times(t-t_0))+s)Sx \qquad (7)$$

'If the body *x* is at the moment t_0 in position *s* and it moves with the constant velocity *v* then, at the moment *t*, the body *x* is in the position $(v\times(t-t_0))+s$'. According to this formulation, the uniform motion law doesn't refer to numbers, but to moving bodies. Numbers are not values of the individual variables but they are components of the expressions of the predicate. Consequently, numbers are not objects, therefore, the mathematical realism, deriving from logical realism, has no foundation.

References:

Beaney M., ed., 1997, *The Frege Reader*, Wiley-Blackwell, London.

Brenner J. E., 2008, *Logic in Reality*, Springer, Dordrecht.

Bynum T. W., ed., 1972, *Conceptual Notation, and Related Articles*, Oxford U.P., Oxford.

Cocchiarella N. B., 2007, *Formal Ontology and Conceptual Realism*, Springer, Dordrecht.

Hochberg H., 2001, *The Positivist and the Ontologist: Bergmann, Carnap and Logical Realism*, Rodopi, Amsterdam, Atlanta.

Lovejoy A. O., 2007, *The Revolt Against Dualism*, Read Books, New York.

Muirhead J. H., 2007, *The Platonic Tradition*, Read Books, New York.

Sorensen R., 2003, *A Brief History of the Paradox*, Oxford U.P., Oxford.

The Meaning of the Logical Constants and Classical Negation

Bogdan A. DICHER*
Babes-Bolyai University Cluj-Napoca

Abstract:
In this paper I review the project of providing of proof-theoretic justification of the logical laws, with a particular emphasis on the possibility of justifying classical negation.

Keywords: logical constants, negation, meaning, proof theoretic semantics.

Anti-realism is a philosophical project whose negative consequences are more or less well-known. In fact, to some extent, anti-realism is received as nothing but a set of negative claims about how classical logic and some of the philosophy accompanying it are mistaken. Yet, this is not all there is to be known about anti-realism — although, one must accept, this fact is sometimes hard to discern not in the least because of the anti-realist way of presenting its central tenets.

This is, for instance, the case of Dummett's assault on realism: most of the ensuing discussion focused on the (semantical) challenges he raised against realism and on their logical revisionary implications, while relatively little attention was paid to the way his positive suggestions concerning the meaning of logical constant fare by his very own criteria. (Which is by no means to say that the 'little attention' mentioned before is not impressive.)

* E-mail: bdicher@gmail.com.

This is the task I set up for this paper, in which I shall (I) review Dummett's account of the meaning of the logical constants together with his development of the concept of *harmony*, and then (II) review some of the criticisms raised against his suggestions regarding the classical rules for negation; also, I will consider the availability of an extended framework for presenting and justifying the classically understood meanings of the logical constants, that of multiple-conclusion logics.

I

1. Philosophers usually think of the meaning of the logical constants as being determined by the familiar truth-tables. In other words, they think that these are to be explained in a model-theoretic style (and, perhaps, that this is *the only* way to do the job). A model-theoretic explanation of this sort basically tells us what happens with the truth value of a sentence, given the truth values of its component sentences. This can be easily seen when we consider the familiar truth-table for, say, conjunction:

p	q	p *and* q
1	1	1
1	0	0
0	1	0
0	0	0

As we can read from the last column of the table, there is only one situation in which the conjunction of *p* and *q* is true: namely, when both conjuncts are true. Is it plausible to think of this as an explanation of the meaning of 'and'? Yes, insofar as it exhibits what must be the case for a sentence formed by means of 'and' to be true. It is another question whether this is all there is to be known about this logical constant in order to make good use of it. Most likely this is not true as there is little of use in what has been said above when it comes to discerning between correct and incorrect

uses (and here I mean actual uses) of 'and'. In other words, if we take logic to be about proper reasoning, the table 'mumbles' rather than 'speaks clearly'. However, most of what is logically relevant for this matter can be straightforwardly explained taking the truth-table as the starting point; in this way, a little gloss can transform the mumble into a clear indication. For instance, the table, at its first line, tells us that, whenever two statements are true, so is their conjunction. As such, it appears to motivate indisputably the familiar rule of 'and'-introduction:

$$\frac{A_1 \qquad A_2}{A_1 \wedge A_2}$$

A parallel reading motivates the elimination rules (where *i* ranges over 1,2):

$$\frac{A_1 \wedge A_2}{A_i}$$

It is obvious that, by the same line of the truth-table, if a conjunction is true, then both its conjuncts are also true, and therefore one is entitled to infer whichever one pleases.

Not all the rules for the (use of) logical constants are so straightforwardly explained in terms of their truth tables. Consider the case of disjunction; according to its truth-table, a disjunction is true when at least one of its disjuncts is true:

p	q	p *or* q
1	1	1
1	0	1
0	1	1
0	0	0

This clearly motivates the familiar rules of introducing 'or' (*i* ranges over 1,2):

$$\frac{A_1 \qquad A_2}{A_1 \wedge A_2}$$

But it fares less well as far as the elimination rule is concerned, for there is nowhere to go in a deduction from the knowledge that one of two sentences is true. (Of course, this knowledge together with the knowledge that one of the disjuncts is not true leads to something, namely, the assertion of the disjunct not known to be false.)[1] The rule of 'or'-elimination is:

$$\frac{A \vee B \qquad \begin{matrix}[A] \\ C\end{matrix} \qquad \begin{matrix}[B] \\ C\end{matrix}}{C}$$

In effect, it tells one that if each disjunct independently leads to the same conclusion, then that conclusion is a consequence of the disjunction. We can use the truth-table to make some sense of this: we could say that, as the first three lines of the truth-table indicate, there is a certain amount of 'indeterminacy' embedded into the meaning of a disjunctive sentence. Therefore, the rule for eliminating it should somehow be responsible to this indeterminacy; the 'natural' way to accomplish this is to 'act as if'. This pretense is captured by the two assumptions A_1 and A_2 and the subsequent subdeductions; its harmless character follows from the fact the assumptions end up to play no role in our assurance that C, so they are discarded (or closed)—fact represented by enclosing them into square brackets.

(The above play with the notion of 'indeterminacy' is even clearer in the case of the rule for eliminating the existential quantifier:

$$\frac{\exists x \Phi(x) \qquad \begin{matrix}[\Phi(a)] \\ C\end{matrix}}{C}$$

The major premise of this instance of the rule of existential elimination tells us that there is an object satisfying condition Φ; it does not, however, tells us which is that object. So, we can suppose that that object bears the name 'a' and if C is a consequence of this assumption, then it is a

[1] I have switched from talking about true sentences to sentences known to be true. Although it is not my intention to do so, in the present case it is entirely harmless: I was after all talking about deduction, and these are man-made, so our knowledge of what goes inside them is relevant.

consequence of the major premise. The usual constraint on such a rule is that the name '*a*'—the parameter—appears in no premise, nor in *C* or in any sentence on which the premises and *C* depend; and the motivation for the constraint is that it warrants that nothing more is assumed about '*a*' other than its satisfaction of Φ.[2])

Apparently, we have managed to read the elimination rule for disjunction from its truth-table. Closer inspection of the previous motivation for 'or'-elimination will reveal that, in fact, we have used a great deal more information than the truth-table comprises: We have paid significant attention to what happens in a deduction if we wish to develop it beyond the occurrence of a disjunctive sentence; we have played with the truth-table-inspired idea that some disjunct must be true, although we do not know which etc., but in doing so we have devised instruments which go way beyond what the truth-table captures, e.g., the idea of making assumptions and discard them subsequently.[3]

3. Fortunately, this is not the only way in which we can provide a reasonable explanation of the meaning of the logical constants. A more straightforward way, at least with respect to their behavior within derivations, is to consider directly the rules governing them. They are all too familiar, thanks to the work of Gerhard Gentzen.[4] His starting idea for the development of the natural deduction presentation of logic was to bridge the gap between actual mathematical reasoning and the formalization of logic in the tradition of Hilbert. Natural deduction, he claimed, 'comes as close as possible to actual reasoning' (CW, 68). Something more was achieved, nonetheless, and to some extent Gentzen seems to have been aware of this.

---

[2] Obviously, there is no question of truth-tables in the case of the rules concerning the existential quantifier, instead, we shall speak of the usual Tarski-style specification of its meaning: an existentially quantified sentence is true just in case there is an object in the domain of interpretation which satisfies it.

[3] A straightforward reading of the or-elimination rule is, nonetheless, possible, provided that we allow our concept of logical consequence to lead to multiple-conclusions. I shall have more to say about this latter on.

[4] He first presented the system of natural deduction in his 1934 paper, 'Untersuchungen über das logische Schlissen' (*Mathematische Zeitschrift*, 39, pp. 176-210; 405-431). Page references are given to the English translation in M. E. Szabo (ed.), *The Collected Works of Gerhard Gentzen* (North Holland: Amsterdam, 1969), hereafter referred to as CW.

He points out that '[t]he introductions represent, as it were, the "definitions" of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions' (CW, 80). In effect, Gentzen suggests that the rules of natural deduction, and the introduction rules in particular, could represent a new way of explaining the meaning of the logical constants.

There is no further development of the point in Gentzen's paper; however, the brief remark quoted above made quite a career in the philosophical thinking of Michael Dummett and Dag Prawitz.

4. The natural unfolding of the idea is easy to foresee: we have a specification of the meaning—or at least of some core part of it—of the logical constants; we need to further develop it in order to explain as many of the relevant aspects of their 'behavior' as possible; and we must make sure that we have grasped the constraints which govern the development. I shall retrace it in the work of Dummett, with occasional, if significant, glimpses in that of Prawitz.

It may be useful to sketch the motives which sustain Dummett's pursuit of this line of thought; the fact that they are well known will justify a sketchy account. (As far as the actual argument is concerned, this stage could be overlooked: if we take, as we should, the construction of a justificationist semantics for the logical constants as the first step in the development of the positive agenda of Dummett, then the success or failure of it could be appraised independently of the reasons that have prompted it.)

Dummett's most general claim is that the Tarski-inspired (hereafter, I shall say 'realist') account of the meaning of the logical constants renders their meanings less than completely transparent to use and this lack of transparency is a capital sin; this is the upshot of the manifestation and acquisition challenge. The realist construal already encompasses an account of meaning (specifically, of the meaning of the constituent statements) which surpasses what can be justified in accordance with the requirement that there be nothing in the meaning of a statement which goes beyond the actual use made of it. This is easy to see in the case of disjunction: in the above model-theoretic attempt to explain it the underlying idea is that a disjunction may be true even if we do not know which of its disjuncts is

true. A step further will also point out that the truth of the disjunct is taken to be independent of us having recognized it as such.

Arguing that a compositional meaning theory renders the logical laws both susceptible and in need of justification, he then points out that the usual semantical (i.e. Tarskian) explanation of the meaning of the logical constants is (pragmatically) circular, and, to that extent, it lacks explanatory power. This is so because, Dummett claims, a Tarski-style explanation of the logical constants appeals (being 'disquotational') to a merely 'programmatic' notion of interpretation which guarantees that whatever logical laws are assumed to hold good in the metalanguage will also hold in the object language. The whole project of providing a justification of the logical laws is thus vitiated.[5]

No form of circularity threatens a proof-theoretic justification of the logical laws. Consider the case of a derived rule of inference: this is sometimes thought of, perhaps not entirely appropriate, as a kind of shortcut. But the whole point of a derived rule is that it provides us with means of establishing some conclusions and our reasons for accepting this is that we have already accepted the initial rules. If the entire justificationist program would be of this sort, then it would be grossly circular. But it is not necessary to be so, provided we are willing to regard, as Gentzen suggested, the introduction rules for the logical constants as primitive in the sense that they stand in no need for justification, or, rather, in the sense that they are self-justificatory. They are not only constitutive of the meaning of the constants, they are the *foundation* of those meanings. Anything more that can be considered to be part of the meaning of the logical constants must be acknowledged to be so only if it accords, in some relevant ways, with the introduction rules.[6]

---

[5] See Michael Dummett, *The Logical Basis of Metaphysics* (Duckworth: London, 1991), 200-204 (hereafter, I shall refer to it as LBM).

[6] Note that there is no pre-theoretical neutral reason (at least, no apparent one) for thinking that it is the introduction rules which are meaning-determinative; this option is sustainable only in the context of 'systematic theory, which will provide for the derivation of all other features of use from that which has been selected as the central notion of the theory' for

5. This requirement is generically formulated by Dummett as the requirement for *harmony*. Briefly, the point is that the two aspects of our logical practice, that of inferring a complex statement from less complex statements, and that of drawing simpler consequences from more complex premises must be in some equilibrium in that whatever warrants the assertion of complex statements is not lost when we move towards their simpler consequences. In Dummett's words and contrapositively:

> The disharmony means that we are accustomed to draw conclusions from statements made by means of **E** [a generic expression] that what we treat as justifying the assertion of those statements does not entitle us to draw. (LBM, 218)

And further,

> Those conclusions (…) cannot consist of statements containing **E**; for the drawing of *such* conclusions must count as part of our convention governing the justification of assertions involving **E**. If there is disharmony, it must manifest itself in consequences not themselves involving the expression **E** but taken by us to follow from the acceptance of a statement **S** containing **E**. (LBM, 218)

The above is a characterization of a form of global harmony, whose formal analogue is the requirement for *conservativeness*, in the particular case that concerns us, for any addition of logical constants to a given language:

> Any one given logical constants, considered as governed by some set of logical laws, will satisfy the criterion for harmony provided that it is never possible, by appeal to those laws, to derive from premises not containing that constant a conclusion not containing it and not attainable from those premises by other laws that we accept. (LBM, 219)

---

only such a theory will allow one 'to pick one or the other type of rule as *the* distinguished determinant of meaning' (LBM, 217).

A more modest, or *local,* understanding of harmony, also at play in Dummett's justificationist semantics, meant to characterize only the introduction and elimination rules for a logical constant, is simply the equation of the requirement with that of normalizability (or, more colorfully expressed, the possibility of 'leveling local peaks').[7]

(There is also a technical motivation for the requirement that the addition of a new logical constant to a language yields a conservative extension. Notoriously, Arthur Prior suggested that the proof-theoretic semantics is meant to fail.[8] He devised a logical constant, 'tonk', whose introduction rules were the rules for 'or'-introduction and whose elimination rules are identical with those of 'and'. Under these conditions, one can prove anything, so the system collapses. The solution to the threat, proposed by Belnap,[9] was exactly the requirement of conservativeness: what went wrong with *tonk* is precisely the fact that while '*A* tonk *B'* has rather lax conditions of assertibility, it also has (unjustifiably) permissive rules of elimination.)

6. Next we can proceed to examine the development of the proof-theoretic or justificationist explanation of the logical constants.

Dummett distinguishes three 'levels' at which it makes sense to speak of a justification of the logical laws. First degree justification, consisting in the 'familiar process of deriving a given law from others' (LBM, 245), reduces the question concerning the justification of one law to that of the justifiability of a previously accepted law. Second degree

---

[7] See LBM, 250. Here Dummett subscribes to Prawitz's suggestion, technically developed in his *Natural Deduction: A Proof Theoretic Study* (Almquist & Wiksell: Stockholm, 1965) (hereafter referred to as ND). The sense in which normalizability can be taken as a form of harmony is evident once we consider the ND formulation of 'the inversion principle': 'Let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition (…) for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already "contain" a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α' (33) (the passage is italicized in the original). That this pertains to more than just the development of normalization is also pointed out by Peter Schroeder-Heister in his 'Validity concepts in proof-theoretic semantics'(*Synthese*, 148, 2006, 525-571), 533.

[8] See his 'The runabout inference ticket' (*Analysis*, 21, 1960, 38-39).

[9] Nuel D. Belnap, 'Tonk, plonk and plink' (*Analysis*, 1962, 22, 130-34).

justification—in effect, the justification of the elimination rules on the basis of the rules for introduction—doesn't rely simply on the assumption that the justificatory laws are valid; rather, one must further claim that this set is 'in some sense' complete. As such, it purports to show that '*any* elimination rule (…) is in harmony with the introduction rules' and, consequently, justifiable and valid (LBM, 253). The upshot of this procedure is to show that

> If a statement whose principal operator is one of the logical constants in question can be established at all, it can be established by an argument ending with one of the stipulated introduction rules. (LBM, 252)

This is in fact the procedure at the hart of Prawitz's proof of the normalization theorem. The point of the theorem is that any occurrence of a formula in a given derivation with the following two features (*i*) it is the result of an application of an introduction rule and (ii) it is followed immediately by an application of the corresponding elimination rule can be removed. So, let *D* be a derivation of this kind for, say, a formula whose principal operator is 'and'. Then, by the two conditions above, *D* has the form:

$$\dfrac{\dfrac{\begin{matrix}\delta_1 \\ \vdots \\ A_1\end{matrix} \quad \begin{matrix}\delta_2 \\ \vdots \\ A_2\end{matrix}}{A_1 \wedge A_2}}{A_i}$$

(The conclusion A_i represents whichever one pleases of the two possible conclusions, A_1 or A_2.)

It is easy to see that the same result could have been obtained by a deduction D^* which has the form of one of the two subdeductions preceding the occurrence of the maximal formula (or local peak) '$A_1 \wedge A_2$'.[10]

Finally, and as a consequence of the fact already hinted at in 1 above that some of the logical rules governing the logical constants are more

[10] For a complete list of the reductions, see ND, 36-38.

complex than this (and similarly) simple case(s), one needs to devise a procedure for a third-degree proof-theoretic justification of the logical laws. Basically, this is a definition of validity for an arbitrary argument, relying solely on its shape. Since Dummett's own description of the procedure is rather cumbersome, it is probably a better idea to begin with Prawitz's simpler treatment, and then to move to a differential description of Dummett's.[11]

I have previously spoken of 'derivations' roughly in the sense of (quasi)formal analogues of regular (i.e. non-formal) arguments. We can now refine this distinction, by adding the specification that a derivation becomes an argument as soon as one provides a justification for the steps it contains. So, I will subsequently use 'derivation' as a substitute for what Prawitz calls 'argument skeleton'. Derivations can be *open*, if they contain undischarged assumptions or free variables, closed otherwise. A derivation is in *canonical* form if it ends with an application of an introduction rule (furthermore, if the derivations of its premises are valid, then the argument will be also valid). A derivation in canonical form will count as *valid* if its closure is valid. (A derivation's closure is the result of substituting a closed derivation for each open assumption in it.) A closed derivation (argument skeleton) is valid if either it is in canonical form, or it can be reduced to a derivation in canonical form (with the same final point).[12]

6. As Prawitz points out (MAP, 517) Dummett's definition diverges from this mainly by making 'the more stringent requirement' that every subdeduction of a given deduction be in *canonical form*; thus, we can speak of *hereditary canonical forms*. This is definable, in the present terminology by adding to the demand for canonicity the extra requirement that 'in case the introduction does not bind any assumption or variable, its immediate subarguments are also in hereditary canonical form' (*ibid*.). This difference is also observed by Dummett, which motivates the introduction of the

---

[11] There are several places in which Prawitz exposes his definition of validity (and, by the same token, his explanation of the meaning of the logical constants); the present account will rely mainly on his paper 'Meaning approached via proofs' (*Synthese*, 148, 2006, 507-524); hereafter, MAP.

[12] MAP, 515.

requirement for heredity by noticing that, unlike Prawitz who is interested in justifying only the elimination rules, his interest is in demonstrating the validity of arbitrary inference rules. Thus, since the procedure may concern a derivation which contains a more complex rule of inference, it may be necessary to proceed by reductions in more than one occasion.[13]

7. The fundamental idea here is what Prawitz presents as the 'inversion principle'; the same point appears in LBM, this time under the name of 'fundamental assumption':

> If we have a valid argument for a complex statement, we can construct a valid argument for it which finishes with an application of the introduction rules governing its principal operator. (254)

The intuitive upshot of this assumption is that if we have whatever means (and we specifically target *indirect* ones) of establishing a proposition as true, then we also posses (at least in principle) *direct* means of establishing it as such. (This formulation is more comprehensive than necessary in the present context, because we are here interested with devising proofs; however, proofs are not the only means of establishing a proposition to be true—at least, not if we are unwilling to regard such things as seeing or listening to the testimony of an eye-witness as means of the sort indicated.) The fundamental assumption was present throughout the entire discussion of the previous paragraph; it is most evident in the formulation of the idea of second degree justification procedure.

The problem with the fundamental assumption is that the proof-theoretic semantics which so heavily depends upon it does not justify classical logic — in effect, it fails to provide a justification for classical

---

[13] See LBM, 255; 252 for the references to Prawitz's work. Dummett's own definition of canonical argument is the following: '(i) its final conclusion is a closed sentence; (ii) all its initial premisses are closed atomic sentences; (iii) every atomic sentence in the main stem is either an initial premiss or is derived by a boundary rule; (iv) every closed complex sentence in the main stem is derived by means of one of the given set of introduction rules' (LBM, 260). A 'boundary rule' is a rule for manipulating atomic sentences.

negation and so it seems to provide another argument in favor of a broadly intuitionistic revision of logic. Let us see why this happens.

8. In order to see just what happens with the attempt to justify proof-theoretically the classical negation, it is useful to take a closer look at Prawitz's proof of the normalization theorem.

From the familiar system of minimal logic, comprising the rules for implication, conjuction, and disjunction we obtain intuitionisitc and classical systems by adding suitable rules for negation. There are several options: a very economical one, adopted by Prawitz, is to enlarge the language by a constant symbol for falsity or absurdity $\perp$. Then, the following two rules will govern, respectively, intuitionistic and classical logic:

$$\frac{\perp}{B} \quad (\textit{ex falso quodlibet}, \text{ for short EFC})$$

$$\begin{array}{c} [\neg A] \\ \vdots \\ \perp \\ \hline B \end{array} \quad (\text{classical } \textit{reductio ad absurdum}, \text{ CRAA})$$

The negation $\neg A$ of A can be then introduced as an abbreviation for $A \rightarrow \perp$. CRAA can be replaced by double negation elimination; or even by the addition of *tertium non datur* to the intuitionistically valid rules.[14]

Unfortunately, either variant spoils the symmetry of the previous rules. This is obvious in the system used by Prawitz: whereas every other

---

[14] This was Gentzen's option. If we take negation as primitive, then the most significant difference is that we would have to formulate EFC so that the absurdity (contradiction) becomes explicit: $\dfrac{A \quad \neg A}{B}$. The same job, in the absence of negation could be done by Peirce's law: $\dfrac{((A \rightarrow B) \rightarrow A) \rightarrow A}{A}$. See A. S. Troelstra, H. Schwichtenberg *Basic Proof Theory* (Cambridge University Press: Cambridge, 1996), 47.

logical rule is either an introduction rule or an elimination rule, the rules for negation obviously do not fall in either category.

One could think that if we take negation as a primitive symbol and use the corresponding form of the (intuitionistic) *reduction ad absurdum* as an introduction rule and the double negation elimination as the elimination rule then, since we recover some form of symmetry, we can reason as in the case of the other rules. However, this is not true, simply because there are proofs—e.g. that of the law of the excluded middle—where there are irreducible maximal occurrences of some formulae. Specifically, the result will be established by applying DNE to what was previously inferred by negation-introduction.[15]

Here is an illustration, in the case of a proof of the law of the excluded middle:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\neg A]_1}{\neg A \vee A} \qquad [\neg(\neg A \vee A)]_2}{\neg\neg A}\,{}_1}{\neg\neg\neg A}}{A} \qquad [\neg(\neg A \vee A)]_2}{\dfrac{\dfrac{A \vee \neg A}{\neg\neg(A \vee \neg A)}\,{}_2}{\neg\neg(A \vee \neg A)}}$$
$$A \vee \neg A$$

This proof contains at least two 'local peaks', in Dummett's terminology (maximal formulae in the usual one) which cannot be removed: there is no way to obtain the desired result without introducing negation and then removing it. It is easy to see that not even the use of the falsity constant can repair this situation. Dummett's conclusion is that 'Plainly, the classical rule is not harmony with the introduction rule' (LMB, 291).

He must mean by this 'local harmony', for it is not clear as yet that it also lacks total harmony, i.e. that it produces a non-conservative extension

[15] ND, 34-35.

of the language to which negation, as governed by the classical rules, is added. But this is easy to show, by providing instances of logical laws which, although themselves not containing negation, cannot be proved without appeal to it. The show-case for this is the so called Peirce's law:

$$\dfrac{\dfrac{\dfrac{\dfrac{[\neg A]_1 \quad [\neg A]_2}{B}}{A \to B} \qquad [(A \to B) \to A]_3}{A}_1}{\dfrac{\dfrac{\neg\neg A}{\neg\neg A}_2}{A}_3}$$

$$(A \to B) \to A) \to A$$

This motivates the charge of global disharmony. Dummett is well aware of the fact that the intuitionistic rules for negation present themselves not so well either. In this case, a local peak would have the form:

$$\dfrac{\dfrac{\begin{array}{c}[A] \\ \vdots\delta_2 \\ \neg A\end{array}}{\neg A} \quad \vdots\delta_1}{A} \quad \text{which is reducible to:} \quad \dfrac{\begin{array}{c}\vdots\delta_1 \\ A\end{array} \quad \vdots\delta_2}{\neg A}$$

$$B \qquad\qquad B$$

This case of reducibility is rather peculiar, because

> The local peak has been lowered, but not leveled, in that we have not found a way of arriving at the final conclusion B from the initial premisses of the argument without the use of the negation operator. (LBM, 292)

The procedure Prawitz envisaged for constructing such reductions consists in actually transforming a deduction containing a maximal formula occurrence into one in which the major premise (of the elimination rule) occurs no more. For instance, in the case of a maximal formula whose main operator is the implication, the derivation to be reduced and the derivation to which the former is reduced are:

$$\begin{array}{ccc} & [A] & \\ & \vdots\delta_2 & \\ \vdots\delta_1 & B & \\ \underline{A \qquad\qquad A \to B} & & \\ B & & \end{array} \quad \text{which reduces to:} \quad \begin{array}{c} \vdots\delta_1 \\ A \\ \vdots\delta_2 \\ B \end{array}$$

We no longer have the major premise occurring in the reductive derivation, while in the former case, of negation, we have not managed to remove its occurrence. The explanation for this is that the rule by which the maximal formula was inferred is not, despite our taking it to be, a genuine (self-justifying) introduction rule. According to Dummett, such a rule would have to be 'single-ended', that is, to work solely as an introduction. RAA fails to comply to this requirement because, as shown by the above (quasi-successful) reduction, the use of the introduction must rely, no matter what, on use of the constant it is supposed to be introducing.

The rehabilitation of (intuitionistic) negation is interesting, as Dummett claims that *ex falso quodlibet* can be used to justify (non-classical) *reduction ad absurdum*; so, instead of seeing the introduction rule as self-justifying, one proceeds the other way round, from the elimination rule to the justification of the introduction. Now, the general form of *ex falso quodlibet* could be taken to be:

$$\frac{G \qquad F}{H}$$

The restricted form being, obviously, obtainable by letting F be $\neg G$. This suggests the following introduction rule (under the assumption that in the above schema the leftmost formula is the major premise):

$$\begin{array}{c} [G] \\ \vdots \\ H \\ \hline F \end{array}$$

In the case of negation the major premise would be $\neg A$, the minor A and the conclusion B. So, the purported introductory rule, having the minor premise

as assumption, the major as conclusion and the conclusion of the elimination rule as consequence of the assumption, would have to be:

$$\begin{array}{c} [A] \\ \vdots \\ \dfrac{B}{\neg A} \end{array}$$

Clearly, this does not work, for it amounts to the claim that our warrant to assert a negated sentence is that its affirmation permits the assertion of a sentence. However, one should note that the rule for negation elimination is such that its conclusion bears no structural relation to its premises. Therefore, the conclusion can be '*any* atomic sentence (supposing the restricted rule to yield an atomic conclusion)'. Similarly, the proper rule for introducing negation must reflect that and it must be infinitary: it is not the fact that *A* leads to the derivation of *B* that warrants the negation of *A*; rather, it is the fact that *A* leads to all (any) statement of the language. This fact is representable by means of RAA, provided we are willing to regard the absurdity constant as the conjunction of all the atoms in the language. Thus, we are taken back to the formulations of *ex falso quodlibet* and RAA in terms of the absurdity constant.[16]

9. Before proceeding, it would be helpful to take stock of what has been said so far:

(i) A Tarski-style explanation of the logical laws is merely programmatic and therefore it is unhelpful;

(ii) it is possible to provide a foundation for our logical practice which is unbiased by model-theoretic notions;

(iii) this attempt succeeds as far as intuitionistic logic is concerned, but it fails to vindicate classical logic; in particular, it show that classical negation cannot be accounted for in this way.

Subsequently, I shall let *(i)* barely touched and I shall assume, for as long as possible, that *(ii)* is not incorrect. Most of the discussion will be concerned with the correctness of *(iii)* (it will become clear that some

[16] LBM, 293-95.

problems it raises can induce significant discomfort for someone disposed to endorse *(ii)*, this is why I called it an 'assumption').

II

10. The above claimed failure of inferentialist semantics to vindicate classical logic depends on the claim that harmony is to be equated with conservativeness. This means that conservativeness is both necessary and sufficient for harmony. That it satisfies the sufficiency condition if fairly obvious: if a given newly added logical constant ensues in a conservative extension of the initial vocabulary, this means that the set of ***k***-free provable statements remains unaltered. (In particular, it is not augmented.) By consequence, the set of reasons on which any such statement depends is unaltered; and, if the previous set of logical constants was harmonious, so will be the one obtained by adding ***k***.

The situation is not as fortunate as far as necessity is concerned. The fact that harmony entails conservativeness is seriously shaken by the following simple observation of Prawitz:

> From Gödel's incompleteness theorem we know that the addition to arithmetic of higher order concepts may lead to an enriched system that is not a conservative extension of the original one in spite of the fact that some of these concepts are governed by rules that must be said to satisfy the requirement of harmony.[17]

The same point the same point can be made by way of a truth-predicate. Consider the case of a predicate T governed by the following two rules:

$$\frac{A}{T(A)} \qquad\qquad \frac{T(A)}{A}$$

[17] Prawitz, 'Review of Michael Dummett, *The Logical Basis of Metaphysics*' (Mind, 103, 373-76) at 375.

The two rules are clearly harmonious; yet, in light of the incompleteness result, T yields a non-conservative extension of the language (or, better, of the theory) in which the Gödel sentence would be provable.[18]

It is difficult to appraise the impact of this failed identification on the significance of the demand for harmony. In particular, it seems that the less demanding concept of local harmony, i.e. normalizability, could work just as well, at least as far as the purported argument against classical negation is concerned. Furthermore, understanding harmony as a demand upon the proper form of an elimination rule could perhaps serve sufficiently well, provided we do not resist some radical changes in the presentation of our logic.[19]

11. I shall now consider a distinct approach to the requirement for harmony, together with a tentative argument showing that classical negation *is* governed by acceptable rules.

Essentially the same points about Dummett's conception of harmony are to be found in Stephen Read's 'Harmony and autonomy in classical logic'[20] (alongside some further developments and criticism of Dummett; I shall return to these latter). Of much more interest here is the positive proposal of Read: that harmony is best understood as a special type of relation between introduction and elimination rule, being implied that there is no need for further constraints once this relation can be shown to obtain. In a sense, Read's proposal is a 're-Gentzenification' of the idea of harmony.

Suppose we have a specific logical constant, $\boldsymbol{k}$, governed by the following introduction rules:

$$\frac{\delta_1}{A_k} \quad \frac{\delta_2}{A_k} \quad \ldots$$

(The subscript indicates that $\boldsymbol{k}$ is the main operator in A.)

---

[18] But see also Göran Sundholm, 'Proofs as Acts and Proofs as Objects: Some questions for Dag Prawitz' (*Theoria*, 64, 2008, 187-216).

[19] See below, the analysis of Read's proposal.

[20] In *Journal of Philosophical Logic*, 29, 2000, 123-154.

The requirement of harmony, claims Read, that whenever something can be inferred from *A* it can also be inferred from what *A* was inferred (here $\square_1$) motivates the following general elimination rule for ***k***:

$$\frac{A \qquad \begin{matrix}[\delta_1] \\ B\end{matrix} \qquad \begin{matrix}[\delta_2] \\ B\end{matrix}}{B}$$

(Where the square brackets surrounding the two subderivations δ_1 and δ_2 mark the fact that whatever assumptions *B* depends on are being discharged by the application of the rule.)

This is easily normalizable: for, if A_k is inferred from the grounds represented in δ_i and immediately subjected to an application of the elimination rule we have that:

$$\frac{\begin{matrix}[\delta_i] \\ A\end{matrix} \qquad \begin{matrix}[\delta_1] \\ B\end{matrix} \qquad \begin{matrix}[\delta_2] \\ B\end{matrix} \quad \ldots}{B}$$

This maximal occurrence of *A* is easily removed by a reduction to $\dfrac{\delta_i}{B}$.

Consequently, Read claims that:

> It is when introduction- and elimination-rules lie in this relationship permitting normalization that there is the harmony which Dummett seeks in the rules, a harmony between the means by which a formula can be established indirectly and the way it can be obtained directly. (131)

And the significance of harmony, Read adds, is that:

> No more is inferred from a formula containing [***k***] than the introduction-rules for [***k***] warrant. In other words, the constant is entirely logical in character, in that the introduction-rules fully specify the meaning, and so the connective is, in the intended sense, autonomous. (*ibid.*)

In particular, it is not the case to assume that harmony somehow brings with it consistency. (Read shows this by developing a one-place connective governed by harmonious rules which, nonetheless, lead to inconsistency. In effect, that logical connective can be characterized as a 'proof-theoretical liar'.)

12. This conception of harmony suffices to motivate intuitionistic logic; in particular, intuitionistic negation is shown to be harmonious. This is unsurprising, as it uses nothing more than the same device used by Prawitz: define negation in terms of a constant for absurdity and impose no specific rule for its introduction; vacuously, we have an argument for the harmonious character of *ex falso quodlibet*.

In change, it fails—perhaps conveniently—to justify classical logic, and, in particular, the laws governing its negation. An impressive retort to this, developed by Read, is that Dummett's (and Prawitz's) case against classical negation depends crucially on the way their choice regarding the *presentation* of the logical laws. Briefly, their case depends on their decision to present natural deduction rules as single-conclusion rules; opting to liberalize the presentation, so as to allow the case that one application of an inference rule leads to multiple-conclusions, would make it that 'all the negation-free theses of classical logic are provable without use of the rules for negation'.[21] In the present context, we can restrict attention to the rules for implication. In the liberalized, multiple-conclusions form, they are:

$$\text{I:}\quad \frac{[A]}{\Gamma, B}\,\Big/\,\Gamma, A \to B \qquad\qquad \text{E:}\quad \frac{\Gamma, A \to B \qquad \Delta, A}{\Gamma, \Delta, B}$$

(Γ, Δ being multisets.)

With these rules, the proof of Peirce's law becomes:

[21] Read, 144; for a complete set of the rules of multiple-conclusion logic see Troelstra and Schwichtenberg *Basic Proof Theory*, 171; also, for a more philosophically sustained development of the calculus, see A. M. Ungar, *Normalization, Cut-Elimination, and the Theory of Proofs* (Stanford: CSLI Publications, 1992).

$$\dfrac{\dfrac{\dfrac{[A]}{A, B}}{A, A \to B} \qquad [(A \to B) \to A]}{\dfrac{A, A}{A}}$$
$$((A \to B) \to A) \to A$$

Where *thin* and *con* stand for, respectively, thinning and contraction and the assumptions are discharged by the relevant application of the introduction rule. In what follows, I shall leave aside the question of the use of these two rules, observing just that they are the natural counterparts of the sequent-calculi structural rules. To this extent, perhaps it makes sense to demand that they be allowed into natural deduction, as devices for manipulating occurrences of formulae in accordance with the relation of logical consequence. Clearly, the calculus would not work without them.

The claim is that multiple-conclusion logics capture the full theory of the conditional, not just the intuitionistic one, as it is the case where the relation of consequence is restricted so that it is only the class of the premises of a rule that can be larger than a singleton.

13. In a sense, there is nothing out of ordinary with this idea: its roots are in the work of Gentzen, in whose sequent-calculus the only difference between classical and intuitionistic logic is that in for the latter but not the former, the succedent must be at most a singleton. The fact that Dummett and Prawitz reject this presentation of logic has to do with there not being any way of understanding the multiple-formulae conclusion of an argument other than as disjunctively connected. (Note that the elimination rule for disjunction is 'from '*A* or *B*' infer *A*, *B*', without there being any way of pointing to either *A* or *B* as holding.)

Dummett considers and explicitly rejects this extension of the relation of logical consequence. The point is not so much that it presupposes that a disjunct can be true without us being able to identify which disjunct holds true and so it begs the question against the constructivist but that our ways of explaining the structure of such a conclusion necessarily involves

appeal to the meaning of disjunction. This is, obviously, a rather undesirable feature for any attempt to explain the meaning of the logical constants. At the same time, Dummett claims, our grasp of the premises of an inference working together, although in a sense can be explainable in terms of conjunction, does not *unavoidably* necessitates appeal to this particular logical constant.[22]

A better point against this extension of our logic would be to say that simply there is nothing in our usual practice that would suggest that we think of an argument as having more than one conclusion. The strength of the remark is perhaps not impressive, given that the most we can reasonably claim of formal derivations is that they approximate actual proofs and that perhaps faithfulness with respect to the presentation of a proof is less important faithfulness about what the proof achieves.[23]

But, if our logical practice exhibits nothing clearly of this sort, does it also lacks anything reasonably approximating it? Some think there is a point at which, while reasoning as we all think we do, we actually come rather close to the basic idea of multiple-conclusions formalizations of deductive logic. Here is a relevant sample, form Ungar:

> An incomplete proof by cases, for example, can reasonably be regarded as an argument with more than one conclusion (and a completed one may contain more than one occurrence of its conclusion). Even if we believe that a constructive proof, by its very nature, can only establish a single conclusion, there is no reason why its formal representation should not be allowed to contain multiple occurrences of that conclusion.[24]

---

[22] LBM, 187; the latter part of the argument is somehow blurry: just why is it that grasping the idea of jointly asserting some sentences—as premises—would somehow be independent of our grasp of conjunction? For all that I can see, it seems to me that whatever we can say to support this claim is hardly different than various ways of using the conjunction in disguised way.

[23] Somehow in the same vein, but a bit more powerful, Beall and Restall note that the main argument favouring multiple-conclusion logics is that it is highly elegant and useful, while the main counterargument is that nobody uses it (J. C. Beall, Greg Restall, *Logical Pluralism* (Clarendon Press: Oxford, 2006), 13-14).

[24] Ungar, 55.

And further:

> From a constructive point of view, it is perhaps better to think of a derivation with more than one conclusion as an unfinished argument which is completed by showing that a particular formula follows from each of its conclusions (ibid., 56)

I can think of little arguments which could be raised against this point which would not wind up being nothing more than a declaration of dislike, so it is best to look for results by considering what is the picture of logical consequence that connects with this liberalization of the relation. However, before doing this, it may be useful to remark the following about Ungar's defense of multiple-conclusion logics: first, that even if there is no reason not to represent proofs so as to allow for multiple occurrences of their conclusion, neither is there any reason to do so. In fact, it seems that this is an undesirable complication, given that, by his own contention, we have sufficiently effective devices for coping with the situation which mostly resembles that pictured in a multiple-conclusions calculus: namely, the proof by cases.

14. Following Arnon Avron[25], I shall consider a formal concept of consequence relation, which will provide the framework for the discussion. By a consequence relation $\Rightarrow$ (defined on a given language L) we shall understand a relation between multisets of formulae which is (*i*) reflexive; (*ii*) transitive (i.e. satisfies cut); and (*iii*) monotonic. (Avron considers monotonicity—under the alternative name of weakening—only as a *further* requirement. While this is quite useful if one is interested in studying formal consequence relations generally, for our purposes we lose nothing but adding it from the beginning.)

Relative to this relation, one can distinguish between two kinds of connectives, internal and external. The internal connectives are those 'that

---

[25] Arnon Avron, 'Simple Consequence Relations' (*Information and Computation*, 92, 105-140, 1991), preprint downloadable from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.9128; hereafter, SCR.

make it possible to transform a given sequent to an equivalent one that has a special required form' (SCR, 11). In turn, the external connective will be defined as those connectives which allow the combination of sequents, preserving information. The internal connectives are:

a) internal disjunction, +, satisfying: $\Gamma \Rightarrow \Delta, A, B$ iff $\Gamma \Rightarrow \Delta, A + B$;

b) internal conjunction, &, satisfying: $\Gamma, A, B \Rightarrow \Delta$ iff $\Gamma, A \,\&\, B \Rightarrow \Delta$;

c) internal implication, '$\triangleright$', satisfying: $\Gamma, A \Rightarrow B, \Delta$ iff $\Gamma \Rightarrow A \triangleright B, \Delta$;

d) internal negation, '-' , satisfying: (1) $A, \Gamma \Rightarrow \Delta$ iff $\Gamma \Rightarrow \Delta, -A$; and (2) $\Gamma \Rightarrow \Delta, A$ iff $-A, \Gamma \Rightarrow \Delta$ (it useful to think of (1) as defining 'left' negation, while (2) defines 'right' negation);

e) internal absurdity, '$\perp$', satisfying: $\Gamma \Rightarrow \Delta$ iff $\Gamma \Rightarrow \Delta, \perp$

The external (combining) connectives are:

f) external conjunction,' $\wedge$', satisfying: $\Gamma \Rightarrow \Delta, A \wedge B$ iff $\Gamma \Rightarrow \Delta, A$ and $\Gamma \Rightarrow \Delta, B$;

g) external disjunction,' $\vee$', satisfying: $A \vee B, \Gamma \Rightarrow \Delta$ iff $A, \Gamma \Rightarrow \Delta$ and $B, \Gamma \Rightarrow \Delta$.

We can introduce a modification in the definition of '$\triangleright$' (internal implication), so that $\triangleright \Delta$ is empty; in this case, we can distinguish between a strong internal implication—expressed by c) above—and a weak internal implication, corresponding to this modulation. Note also that since we have considered monotonicity as a feature of the consequence relation, there is no distinction between internal and external disjunction. In general, since we shall be concerned only with classical and intuitionistic logic, there is no need—indeed, no ground—to distinguish between the two classes of connectives. I only mention them in order to be permit a characterization of classical and intuitionistic logic be reference to the admissibility of various types of connectives.

All the above conditions can be converted into rules of the familiar sequent calculus. Consider the case of the implication: (c) can be translated as the following set of rules:

$$\text{(R)}\ \frac{\Gamma, A \Rightarrow B, \Delta}{\Gamma \Rightarrow A \rightarrow B, \Delta} \quad \text{(L)}\ \frac{\Gamma \Rightarrow A \rightarrow B, \Delta}{\Gamma, A \Rightarrow B, \Delta}$$

(R), already 'justified' by clause (c), is a perfect match of the right rule for implication in the sequent calculus. The question is whether we can we obtain its left pair? By reflexivity, we have that $A \to B \Rightarrow A \to B$; applying (L) we obtain: (1) $A \to B, A \Rightarrow B$. The premises of the left implication rule in sequent calculus are (2) $\Gamma \Rightarrow A, \Delta$ and (3) $B, \Phi \Rightarrow \Psi$. By applying cut to (1) and (2) we obtain (4) $\Gamma, A \to B \Rightarrow B, \Delta$; with yet another cut applied to (3) and (4), we have: $\Gamma, \Phi, A \to B \Rightarrow \Psi, \Delta$, that is, the conclusion of the left rule for implication. Similar arguments permit the obtaining of the familiar rules for the other connectives; there is no need for an argument in the case of negation, as the simple transposition of the clause in rule-form yields the right and left rules of the sequent calculus. With this, we have a straightforward way of formulating corresponding rules in the framework of natural deduction (with multiple conclusions). Similarly, it is trivial to restrict the above rules to the case in which the considered consequence relation allows but a single formula occurrence on the right hand side of $\Rightarrow$.

15. This view permits a rather simple characterization of the relation of logical consequence which is peculiar to classical logic: this logic makes no distinction between internal and external connectives, and furthermore, it contains all of them. How are we to characterize the intuitionistic consequence relation? Avron considers two distinct possibilities. In the first case, observing the familiar difference between intuitionistic and classical sequent-calculi, one could say the intuitionistic consequence relation is the restriction of the general consequence relation to being single-conclusioned and having external conjunction and disjunction, as well as internal *weak* implication, and absurdity. The second case involves no restriction of the number of conclusions; instead we can impose the condition that Δ (understood as the multiset $B_1, B_2, \ldots$) is a consequence of Γ if and only if the disjunctive closure of Δ is also one of its consequences. This is a simple and philosophically moot equivalence, but it permits exhibiting a nice symmetry between the two calculi. In this case, the difference between the two types of logical consequence is that for the classical one it is required that the implication be strong, while in the intuitionistic case, it can be at most weak. So we apparently have a confirmation Read's contention that by

restricting attention to single-conclusion forms of presenting the rules of natural deduction, Dummett in fact uses a theory of the conditional which is warrantedly weak enough to sustain only the intuitionistic consequence relation. Nonetheless, Avron observes that this presentation of the situation in fact hides 'another crucial difference between the two logics', namely that '*Intuitionistic logic does not contain any internal negation*'. The fact that a given logic has an internal negation is tantamount to the fact that there is an atomic sentence A so that both $A, \neg A \Rightarrow$ and $\Rightarrow A, \neg A$; but this cannot be the case of intuitionistic logic. Just consider what must be case for the latter to be indeed valid—the latter being, of course, the sequent calculus variant of the law of the excluded middle. Now our question is actually how could we prove a sequent with empty antecedent and whose consequent consists of a formula and its negation. (According to the above condition, for this to make sense from the intuitionistic standpoint we must think of it as actually consisting of their disjunction.) Since A is atomic, there is no sequent $\Rightarrow A$, because this would render the system inconsistent. If this is so, than $\Rightarrow \neg A$ must be valid. At the same time, since we assume that no contradiction can be true, $A, \neg A \Rightarrow$ is valid. By applying *Cut* to these two formulas, we obtain: $A \Rightarrow$, i.e. A is absurd.

16. Elsewhere[26], Avron took this to show that intuitionistic negation is not a genuine negation. Of course, this is so under the assumption that the only rules which can govern negation are those previously presented—that is, his point is made taking for granted whatever it is that underlies our conception of classical negation, without paying any attention whatsoever to the stringencies of the semantical theory underlying it. Furthermore, the point can be sustained only if we refuse to accept that the absurdity constant is appropriate to express a form of negation—not the classical one, but another, weaker one. And I see no reason why an intuitionist should be unsatisfied with a perpetual understanding of negation as implication of absurdity. If in turn this is to be understood as the conjunction of all atoms

---

[26] 'Negation: Two Points of View' in Dov M. Gabbay and Heinrich Wansing, *What is negation?* (Kluwer: Dordrecht, 1999, 3-22) hereafter, Negation.

in the language, then indeed we may have a problem. Namely, we may find that intuitionistic negation is unstable: for some languages (including our actual language), the meaning of the absurdity constant will be, indeed, so that it never the case that it is true; for others, i.e., those in which the atoms are mutually consistent, it will be something else. But just how serious is this problem? In other words, what are the prospects for a language which lacks a relation of contrariety various predicates?[27]

This point aside, is there anything helpful in this analysis? I believe it is, namely, I believe that the relation it suggests between a multiple-conclusions presentation of logic and the rules for negation is both stronger and less innocent than Read takes is to be.

The fact is that any relation of logical consequence which has both a strong internal implication and an internal absurdity also has an internal negation, defined, as usually, as an implication of absurdity. In other words, once one accepts both strong implication and absurdity, one has no choice but to admit that negation is indeed regressive, that is, that the double negation of a sentence has as a consequence that very sentence; this is just the half of the double negation which the intuitionists reject. Indeed, there is an important connection between classical negation and multiple-conclusion logics

> Among the various connectives […] only negation essentially demands the use of multiple-conclusions [consequence relations] (even the existence of an internal disjunction does not force multiple-conclusions, although its existence is trivial otherwise). Moreover, its existence creates full symmetry between the two sides of the turnstile. Thus in its presence, closure under any of the structural rules on one side entails closure under the same rule on the other, the existence of any of the binary internal connectives defined above implies the existence of the rest, and the same is true for the combining connectives. (Negation, 6)

---

[27] This worry is voiced by, e.g., Michael Hand, 'Antirealism and Falsity', in Gabbay and Wansing, *What is negation?*, 185-198. In the same collection, Tennant invokes the contrariety relation as a possible solution (see his 'Negation, Absurdity and Contrariety', 199-222). I, however, do not endorse the particular details of Tennant's solution.

Now the point of these observations can be put as follows: even if our main reason, at least according to Ungar, to formalize our logic in a multiple-conclusions form comes from the structure of proofs involving disjunctions, it appears that the main gain of the move is a more powerful theory of the conditional. Even this is not quite obvious once we ask ourselves what is it that makes a classical conditional stronger than an intuitionistic one. The simple answer is that the supplementary strength of the the classical conditional derives from its being sufficient for the construction of proofs which otherwise would require the use of classical negation. But if the latter is unjustified, why would the strengthened conditional be acceptable? Perhaps a good case could be made that this is simply an occult way of introducing classical negation or, better, the meaning-theoretic commitments it carries, into our logic. Granted, we may not be able to argue against it on a purely proof-theoretic basis, but there would still be sufficient meaning theoretic reasons to reject both options.

Because a stronger conditional inevitably brings with it a regressive negation, some might be tempted to claim, as Read did with respect to the weaker (presentation of the) conditional, that this begs the question against the intuitionist. This, however, would be a mistake: for all we can know, this might be what it takes to represent our reasoning according to the laws of classical logic. However, I think that once one rejects this charge, one makes visible the deeper reasons why opting for a multiple-conclusions framework is vexatious.

We must go back to the question concerning the reasons one may have for wanting the relation of logical consequence to be multiple-conclusions. Most likely, it lacks the credentials derivable from our recognition of some form of intuitive logical construction as having the same structure as that exhibited by it. Moreover, consider the claim that a multiple-conclusions framework strengthens the theory of the conditional. This is obviously taken to mean that in this framework the rules of implication allow the derivation of more sentences than it was the case in the single-conclusion framework. Besides this, there is nothing of significance obtained. We clearly don’t obtain a clearer grasp of the

behavior of conditional statements; nor, in fact, can we make sense of the strong conditional except by reference, be it concealed, to the weak one. We can ask ourselves what is the cost at which this was gained? Lesser intuitiveness for the formulation of the logical rules: this is the answer. Is this an affordable price? Yes, if the following condition is fulfilled: that, supposing that we are presented with a derivation having several conclusions, we could recognize it as a derivation without appeal to the weaker conception or without appeal to a previously grasped meaning of at least one logical constant—the disjunction. This, I think, cannot be done.

Needless to say, it would be an indecent display of silly over-optimism to take this as a final stroke in a fight against the multiple-conclusion revolt. If one fancies it, then little can be said to deter one from preaching in its favour; it is another issue whether one can take the reforms suggested to be advisable. For all I can see, I am quite convinced that the reform is needed for a coherent presentation of classical logic. But I also think that this is yet another reason to be cautious about classical forms of reasoning.

17. It is safe to end this paper with a word of caution. To whatever *caveats* I have already expressed, I would like to add the following two. First, although I have tried to make the idea *plausible*, little has been said here in defense of a meaning theoretic explanation of the logical constants. While I think this to be a fruitful investigation route, I am also convinced defending it in its last details is a difficult undertaking. Second, neither the picture of the multiple-conclusions calculus nor that of the classical negation I have here presented is complete. Therefore, if I have indeed managed to cast some doubts on their adequacy, it should be noted that I have merely considered a part of the case in their favor. In particular, some (broadly rejectionists) accounts of negation are worthy of significant attention.

MISCELLANEA

Feyerabend on Fire: Analysis and Critique of Three Arguments

Julian Roel GONZALES [*]
Colorado State University

Abstract:

Paul Feyerabend offers arguments in favor of Democratic Relativism in "Democracy, Elitism, and Scientific Method" that may provide a measure in how we look at science. There are problems in the consistency of his arguments that provide dilemmas in how to implement the changes he wishes to make in a free society, with concern to the scientific view. In a generous analysis of his work, I am at showing how he does not add any sort of new method to understanding science, or its relation to the concerns of the public.

Keywords: Paul Feyerabend, Science, Ideology, Traditions, Democratic, Relativism, Free Society

In writing this paper I hope to show some implications of the three arguments from Paul Feyerabend's "Democracy, Elitism, and Scientific Method" stemming from his belief that a free society is "a society in which all traditions have equal rights and equal access to the centres of power".[1] The arguments he makes are: (1) People have the right to live as they wish; (2) A society that contains different traditions neighboring each other provides a merit based judgment that a monistic society does not; and (3) A science point of view is incomplete due to their lack of significant

[*] E-mail: julianrgonzalez@sbcglobal.net.

[1] Paul Feyerabend, *Science in a Free Society*, Fourth Impression, (London: Verso Editions/NLB, 1987), 9.

phenomena and erroneous in competency. [2] The implications will be shown in a manner of examples that qualify the statements against science he claims, and also some of the problems that his own argument has if followed.

To setup the ways in which we value science we may look at our own grade school education. Learning about the human body and diseases we may catch from a health class, provided a nominal understanding of the body we live in. Most likely our parent's had little knowledge of what we learned, but had the permission slip as consent to teach us "controversial" science. Playing scientist in school laboratories under the supervision of our teachers, gave us an idea of the sort of procedures that scientist go through. The lessons we were taught in these classrooms were valued as much as algebra and history in the eyes of the legislator. Coming from Texas we were always preparing for a standardized test, failing to learn science in the standard they see as fit was penelaized. Science was valued so highly that it was mandatory for our entrance into the next grade. Our education in science prompted us to believe that the status quo was right. It science was compulsory to our lives and any opposition to it was a stance against education, not science. My aim in this section was to set you in the frame of mind which most individuals live by. Science reports given from the media, and probably forwarded emails, are the most interaction with science that individuals have, we have been indoctrinated with and taught not to question.

In an effort to call for a better science some sort of examination into how democracy may be viewed, and a reexamination of what we truly value within science may be needed. It is likely the case that most scientists are not concern with such examinations, so it is up to the public to make inquires into how we do science and how useful it is. If it is the case that scientific thought is dogmatic in training scientist to be averse to "ethical and conceptual" inquiries, then it can be assured that there will be

---

[2] Paul Feyerabend, "Democracy, Elitism, and Scientific Method," from *Inquiry*, vol. 23, 15-16.

malpractice in science due to its distance from common sense.[3] The scientists are set in their ways, so possible means to change if the free society feels it is needed is by investigating and legislating what ought to be changed. This is hard since it has been established that we are trained to think of science as a necessary subject versus an elective.

Free Society/Democracy

The notion of free society he has in mind is unusual from democratic society that we normally adhere to. The way in which we may see the difference is where the rights are placed, commonly we see it as pertaining to the individual who is valued in so far as they have traits that society believes are "constitutive" of their *human nature* and worth protecting to some extent.[4] Where rights are our interests protected by legal and moral barriers, and are guarded from forceful opposition in society.[5] For Feyerabend, it is tradition that is given rights, and is useful. In his definition of a free society it compliments his argument that a pluralistic society, with many traditions, is beneficial in knowing what alternatives are out there to be used. Tradition is given rights because it provides significant meaning to the individual. It is Valued because it is the object that makes life worth living.[6] However, in *Science in a Free Society* he provides a significantly traditional notion of democracy, where it is an assembly of "mature" people and not a collection of innocent followers.[7] The informed learn by their engagement with society's decisions and policy making. Feyerabend values that what he believes to be "maturity" over "special knowledge"; so, scientists believe themselves to be practicing the most important function in society, but society actually puts into play what is useful and trusted in relation to other knowledge.[8] I do not believe that his notion of democracy here is in opposition to his notion of a free society stated earlier. The claim

---

[3] Bernard Rollin, *Science and Ethics*, (New York: Cambridge University Press, 2006), 97.
[4] Ibid., 63.
[5] Ibid., 63-64.
[6] Feyerabend, *Free Society*, 9.
[7] Ibid., 87.
[8] Ibid.

made here takes the notion of how we act in democracy, and compliments the combination of the first argument—people live the way they want—and the second argument—various traditions provide a means to judge best. The combination allows for us to decide in a society what products of science would satisfy and compliment our lives. We are provided with the means to judge it in comparison to our traditions and other's traditions. A trivial example, imagine the judgment and comparison that went into the first microwaves used by the public, and cooks had to decide if it was useful for their lives.

First Argument

Feyerabend argues: "People have the right to live as they see fit".[9] Institutions ought not to be permitted to coerce a Jehovah Witness to take a blood transfusion, since it goes against their tradition. It is the case then that the science becomes a commodity, and people from various traditions can choose to believe the ideas of science. Thus, Feyerabend believes, scientists do not work in judging what is Truth or Falsehood, instead they are salesmen. In this sense the credibility of science is almost a *façade* since it can only be looked at in terms of how it compliments the life one lives.

Another minor example, to show this case, is the debate on whether an egg is healthy or not. We have heard that the egg was good, but then it was bad, and then it was the egg whites that was good and the yolk bad. The progression in research allows for such changes and debunking of old beliefs. However, as a public spectator of science we make the changes in belief analogous to a woman who cannot decide what to wear on a date. We still eat an egg, and disregard the science about what is healthy or not. It is only useful to the extent that our lives are affected by such data. You currently see this science to life relationship in the hysteria over Swine Flu, the first thought a person has is that they should not eat pork. This is how most live their lives in relation to science, if science says something is deadly in a catastrophic sense then we take the advice. It seems as though

[9] Feyerabend, "Scientific Method," 15.

science is most effective in having the public's attention, and trust, when science scares.

The practical applications of science aim at becoming significant to our lives, so that science gives us a sense of control over nature and manipulation for convenience. However, pure science (i.e. string theory), is still needed in order to make that leap into practical. For instance, a physicist running experiments on vortex turbulence in fluid dynamics, can get funding from the military to build a facility and instruments used in experimenting. The inspiration of the scientists in the theoretical is only supported by the hopes of the military for practical uses.[10] It does seem to be the case that as much as science wants to seem to be universal, or can be intrinsically valued—valued in itself, we will only use it as long as it accommodates our way of living.

In valuing science in such a manner, where we will live the way we want and science fits in as long as it does not oppose our tradition, then science is degraded in a way. This seems to be problematic, if we want individuals to be able to look at science and make judgments and policy changes to better science. The two conflicting views that Feyerabend seems to want to endorse are that we ought to live the way we want, and science should be judged by us in how it fits into our lives. If you already look at science in the manners I have describe, as a mere spectator, then you are not getting an insightful view of how science is done, yet he proposes that acting in a democracy (i.e. doing your civil duty) you can create change needed to science.[11] I can accept this, but when you add in the component that makes science a commodity to sell to individuals, then you are valuing science at even a lower level, so you look at science as being nothing more than news reports.

A possible resolution may be to look at science in a way to educate oneself in issues that would pertain to your lifestyle, so if you were a bodybuilder then a men's health magazine may be the special science you

---

[10] Richard Harvey Brown, *Toward a Democratic Science: Scientific Narration and Civic Communication*, (New Haven, CT: Yale University Press, 1998), 136.
[11] Feyerabend, *Free Society*, 87.

need in your life. It happens in philosophy, if you are an applied ethics researcher you look at the journals that cater to that sort of scholarship. It may be the case that you peruse the journals on philosophy of language, but this is no different than the bodybuilder flipping through the pages of *Popular Mechanic*. Fascination is an incredible tool towards learning about science; I think that alone sparks the interest of young students to think about becoming doctors, and biologist. The fascination answered a lot of questions when we were younger. Of course, we become fixed in our traditions and seek science when it is convenient to our lives. Science is readily paid attention to when it scares us, or is demonized. You look at the issue of stem cell research, and how certain religious groups protest science. Making it a moral issue when, the real issue is the definition of a life, and where it begins. I will say that advancements in science do come from a certain general support, so human cloning is not done, and cancer research supported by fellow Texan Lance Armstrong and bracelets. This exemplifies Feyerabend's belief that we view science as a commodity, a fashionable item to our traditions.

Second Argument

By avoiding a monistic tradition society we are able to judge what properties of a tradition we may desire.[12] By judging Feyerabend means in a very specific manner; traditions cannot be judged as "good nor bad," rather they just are. Traditions are comprised of wanted or unwanted properties by those outside of a particular tradition.[13] This is in accordance with his advocacy of a democratic relativism, proliferation in the sense Mill desires: "At such times [periods of transition] people of any mental activity, having given up their old beliefs, and not feeling quite sure that those they still retain can stand unmodified, listen eagerly to new opinions".[14] Thus, people view science, and other traditions, with the ability to be convinced that there

---

[12] Feyerabend, "Scientific Method," 15.
[13] Feyerabend, *Free Society*, 81.
[14] John Stuart Mill, 'Autobiography', *Essential Works of John Stuart Mill*, ed. By Max Lerner (New York: Knopf, 1965), 149.

are some redeeming qualities about a tradition that is not shared in their own. Feyerabend postulates that the reason why we keep traditions, even if they are considered obsolete is because they are still pleasant and they help us comprehend and inspect the most advanced theories.[15] Yet, we are given the means to advance our own traditions and a "maturity" about our acceptance of various cultures in a free society.

There are two occasions in my mind where traditions pertaining to the palate had to be tested, and maybe other bachelors have had this same experience. The occasions involved dates where at one I was offered hummus (of which I thought was haggis) and the other raw fish (sushi, but merely viewed as raw fish). Not to seem as though I were a philistine, and wanted to impress, I gladly accepted the dishes and to my surprise found them delectable. I was matured by the dates, in reference to the food not the women, and given new options in my diet. Food is a good example of how proliferation in traditions allows for us to have a diverse food source and not starve over limited resources. By having options we are given a means to even look at our own traditions and see what really does fit into it and what does not. Prior to sushi I did not believe I liked fish, but in experimenting I discovered its flavor. We already have the right to choose, and this right is exercised in our choice to venture out into the world looking at traditions that we see as foreign.

Let us return to Feyerabend's concern with science superiority complex over our traditions. Science purports that foreign traditions are not to be taken seriously because they do not produce "results" as Western science, technology, medicine, and institutions can.[16] There is no proof, he insist, that supports such a claim that results are better than the problems alleviated by foreign tradition. Secondly, the body is viewed as a machine in Western tradition and thought of in a materialistic sense; where as, other forms of medicine and science see the person—feelings, emotions, and welfare.[17] It is assumed that science is valuable because it can fix, which

---

[15] Feyerabend, "Scientific Method," 7.
[16] Ibid., 13.
[17] Ibid.

seems awfully different from healing. So, medicine has been transformed in its method of helping a person due to molecular biology and sophisticated biochemistry, where "disease was increasingly seen as defects in the machine, and subjective states…'ghosts in the machine'".[18] There is no valuing of the individual in such a method, only the disease or ailment.

There is a discrepancy within the advocacy of this second argument, it seems as though democratic relativism would still have to be valued over the first argument. My reasoning is this: The tradition that goes I do not want any diversity of tradition will have to be disregarded in favor of a proliferation of traditions. Feyerabend addresses this mildly by stating the following: "The belief that the institutions of a free society should protect the individual and not traditions…is correct to a certain extent…it is an incorrect assumption that preservation of these possibilities [rich and rewarding life, or traditions] is a basic value never to be overruled".[19] He does not argue for why preservation of the possibilities can be overruled. If he is correct in presuming this then it seems to bury the first argument because a person may see it fit to live in a manner where other traditions are not accepted, or examined. He seems to mistakenly presume that both the first and second argument are in accordance, and does not leave room for the opposition of proliferation (of whom he should accept as a possible tradition).

Third Argument

In addition to being incomplete, scientific views neglect vital facts and are flawed in their understanding. Resulting in a perpetual assumption in the arguments and procedures, that eventually make the research conducted false or absurd.[20] The scientific view takes for granted the facts that are there to be had, and merely looks at facts that are needed in order to progress the particular science. The progress of science is often at the

---

[18] Rollin, *Science and Ethics*, 218.
[19] Feyerabend, "Scientific Method," 14.
[20] Feyerabend, "Scientific Method," 16.

detriment of the possibility for alternatives; Feyerabend states the severity of this claim in the following:

> Every piece of knowledge contains valuable ingredients side by side with ideas that prevent the discovery of new things. Such ideas are not simply errors. They are necessary for research: progress in one direction cannot be achieved without blocking progress in another. But research in that 'other' direction may reveal that the 'progress' achieved so far is but a chimera. It may seriously undermine the authority of the field as a whole. Thus science needs both a *narrowmindedness* that puts obstacles in the path of an unchained curiosity and the *ignorance* that either disregards the obstacles, or is incapable of perceiving them.[21]

So, it seems as though science builds into it a conflicting scheme, where opposing views are valued for competition's sake, and alternatives ought to be disregarded if there is to be any advancement and credibility in a particular science. Science will not truly acknowledge what the other options are and as a result of the omission it holds itself to a standard that limits its potentiality. By this I follow the second argument, in believing that various traditions can lead to a judgment of what features would best fit their own traditions. In this case I would make a judgment for the sake of a particular science, just like a chef who takes lessons from other cooking traditions in the hope of advancing their own cuisine. Out of the synthesis of traditions it may be capable of advancing the science by combining different traditions.

Feyerabend's argument here seems plausible, considering how we refuse to acknowledge other beliefs if they may take away from the one we hold. This happens in various manners and occurrences, but an example in science can be seen in the misreading of Lloyd Morgan's work on psychology, and the assumption taken by psychologists who have been taught the erroneous view of his work. Morgan's Canon is often taught to be in opposition to the Darwin-Romanes view—that animals have

---

[21] Feyerabend, *Free Society*, 89.

consciousness; yet, in his actual writings Morgan states that there is a consciousness in animals, it is seen in their performance of actions.[22] The psychologists assume that it is correct because the existence of such a belief makes their view and experiments permissible, and as long as they take it as true their view is still correct. However, if an actual assessment of the material occurred it would lead to discrediting their science and making their tradition's advancements false in some cases.

At the heart of this argument science views that it is their job to provide results and that the debates on theories are for those who do not practice the research tradition.[23] They tend to wash their hands of the problems, and it may stem from the ideology that science is value free.[24] Since it is usually taught that science does not make value or ethical judgments, then such debates over the research are not the job of scientist. They do not see these two items as important to their work, and yet the public finds them to be part of science. It seems to be no different than the concern for wanting a particular grocery store to "go green" and be environmentally conscious, the public wants scientists to be concerned for the issues that are not scientific.

Feyerabend's argument essentially advocates for democratic relativism that would provide options, such as medical treatment.[25] By the comparisons of the treatments available an individual can see which option best suits their needs. This would provide better science in that the public would advocate, and use, those particular traditions that pleased their aliments and their personal concerns. This is a shift towards a sort of society that holds science responsible to values and ethical issues, in order to ensure that there will be the best possible science done in the most humane way.

The issue with this argument's conclusion is that there seems to be little done in the actual role of the scientist. They will still be held accountable, as they have been before. The people seek to implement certain

---

[22] Bernard Rollin, *The Unheeded Cry*, (Ames, Iowa: Iowa State University Press, 1998), 75-79.
[23] Feyerabend, "Scientific Method," 16.
[24] Rollin, *Science and Ethics*, 11-30.
[25] Feyerabend, "Scientific Method," 17.

policies by voting in their society. He states it himself that "Everybody knows that cancer research absorbs huge amounts of money" and with little results.[26] Still, we as a democratic society make it known that we want this sort of research done. We are concerned with this, regardless if the scientist want to play in the theoretical aspect, Feyerabend believes this is the reason for the lack of results in cancer research, but this theoretical comes with the benefit of advancing the understanding of cancer. It has to be accepted that the scientist will play the role that we allow them to play; they work for us in a sense because we fund their experiments.

Ultimately, scientists will still keep to their assumptions because they need not venture out into other traditions to show that their tradition works, it is an extra step and unnecessary for them. However, it is the case that society has a certain say in the matter, Feyerabend is correct here, but not in changing the scientists' lack of participation in the debates surrounding science. The citizen can change the direction and drive research, but this is their power over the science and not the scientists' tradition.

Final Concern

The last concern, with his three arguments, is that it is debatable whether they are adding anything to the society that we already function in. People do live their lives the way they see fit, there are a number of traditions available for the public's use, and we already take a stand in how research will be done to ensure science is concerned with important facts. Our legislation and participation in the democratic process has made these possibilities actualized.

The three arguments presented seem to be merely a philosophical review of the established status quo. His advocacy for a democratic relativism by the three arguments is established in our free society, whether he believes it or not. If I want acupuncture I can attain it to alleviate my back pain.

---

[26] Ibid.

In this analysis I hope that there are positive consequences to the measures Feyerabend wishes to take to make science better and a contributing member to our society. Additionally, there are problems how the arguments work in conjunction with one another by the way he presumes they work together. They are conflicting in instances where value of a person's right is overridden by advocacy for proliferation. I have the intuition to believe that there is not much he can say about the changes needed to be made in society, other than that society ought to investigate more and legislate change to science as they see fit.

Collectives as Theoretical Entities

David BOTTING*

Abstract:

The question is often asked whether a group of agents cooperating together constitutes an agent in its own right. I want to approach the problem by starting from a slightly different question: does a group constitute an entity in its own right? From positivism I offer the answer that groups and individual agents are on the same footing with regards to being counted as entities, and from entity realism I add that terms referring to these entities do genuinely refer provided that we can manipulate these entities. There is still a significant difference, though, between individual agents and groups that should not lead us to abandon methodological individualism. A group cannot be an agent in its own right because it does not possess intentional properties in its own right. Individuals are irreducible in a sense that groups cannot be, because no proper part of an individual has intentional properties. Groups are reducible, because they have no properties that cannot be reduced to the properties of its individual members; the group mereologically supervenes, in possibly complex ways, on its members.

Keywords: methodological individualism, collective autonomy, entity realism, reductionism, positivism, intentional stance.

Introduction

The analysis of human action normally starts from the point of view of the individual agent. The individual has a pro-attitude, forms an action-plan to change the world so as to fit that pro-attitude, and acts according to that plan. This is intentional action in its fully-fledged sense. If a collective

* E-mail: davidbotting33@yahoo.co.uk.

qualifies as an agent in its own right, then our account of collective agency can mirror this account of individual agency, and a collectively rational action is performed intentionally when it accords with an action-plan. Does this mean that we need a 'group agent' of some description that *has*, in some sense of *having*, this action-plan? What can this *having* be, since it could not mean *thinking* of an action-plan unless you suppose group minds to exist in some literal rather than metaphorical sense?

Clearly, though, we do need to capture the experience of acting in our daily lives where we have to take other agents into account. The manner of this "taking into account" is not an all-or-nothing affair, and there are different degrees of cooperation, from the simplest cases of considering other agents as mobile obstacles to be negotiated when walking down the road, to agents who share in some sense the pro-attitude and perform acts that, when combined with those of others, aim at bringing the approved state of affairs about.

In this paper I will be concerned with answering the question: what are the relations between individual agents and the collectives of which they are members? The answer will be given: *mereological supervenience*. I will also answer the questions:

A. Do collective actions mereologically supervene on individual actions?

B. Do collective intentions mereologically supervene on individual intentions?

C. Do collective agents mereologically supervene on individual agents?

If the answers to all these questions are "Yes", then we can say everything that we want to say about collective agency by only mentioning the agency of individuals. This is methodological individualism and is marked linguistically by a distributive analysis of ascriptions of agential properties to collectives.

It is beyond the scope of this paper to consider the intricacies of such analysis.[1] What I hope to do instead is a critical survey of the various

1. I have provided such an account in my paper "The Weak Collective Agential Autonomy Thesis". Distributive analysis will be sometimes mentioned in what follows, but I do not want to go into detail about linguistic questions.

accounts of cooperation in the literature and show that none of these require abandonment of methodological individualism. In so doing, I will be arguing that collectives can be considered as theoretical entities albeit not fundamental entities – they are macro-level entities that supervene upon more basic micro-level entities.

Richard von Mises (1968, 229) remarks that there is no obstacle to considering collectives as theoretical entities – from his positivistic point of view, this only puts them into exactly the same bracket as our egos and physical bodies, namely as logical constructs of sense-data – but that this is independent of the truth or falsity of the thesis that social facts are reducible to observation statements that refer solely to individuals. I hold to a principle that we should only license collectivism if different kinds of properties, events, etc., are required at the collective (macro-)level than are required at the individual (micro-)level. To put it another way, once properties of a certain kind have emerged at one level, then exemplification of properties of the same kind at higher levels can always be reduced to exemplifications at that lower level. Agential properties emerge at the level of individual agents – embodied minds – and we should bear in mind that there is an important difference between being an irreducible (in this sense) theoretical entity and a reducible theoretical entity, between individual agents and group agents; this difference grounds the claim that we should endorse methodological individualism. It does not matter if individuals are reducible in the slightly different sense characterized by the naturalistic programme of reducing these agential properties to non-agential properties at even more microscopic levels such as those studied by the physical sciences. I construe the semantics of theoretical claims realistically, subscribing to Hacking's dictum that "If you can spray it, then it's real".

How are collective actions related to individual actions?

Where there is cooperation, means-end reasoning posits acts of which some are to be performed by an agent other than the reasoner. These posited acts may vary in terms of their intentionality and in many cases it is

not part of the goal that an agent has an intention to achieve this goal even though his act is instrumental in performing it; it is enough that the agent perform the act and unnecessary that the agent be free with regard to that act. For instance, a prison gang breaking rocks could be said to act collectively, but this would be an extremely minimal case of cooperation. Another example might be the performance by a battalion of some military manoeuvre when it is not required of the soldiers that they be aware either of the acts of the other soldiers or of the collective action that they are contributing to. They do not share a goal; the conformance of their acts to the specified plan is brought about by military discipline, so the intention has the content of following orders and it is only happy accident, as far as they are concerned, that these orders add up to the successful performance of the collective action. These are cases of heteronomy, where the choice of the goal and the means of achieving it are imposed upon the collective by physical or psychological force. We can call this *heteronomous cooperation*[2] because there is only one person, the battalion commander, who is reasoning. We can still say that the battalion performed the manoeuvre, since it is no part of the act-type performing-a-manoeuvre that its constituent acts (or more strictly, the act-tokens in virtue of which the collective action supervenes on the act-trees of the individual soldiers – more on this later) be performed intentionally.

Let us look at this from the point of view of the Philosophy of Science. I would now like to provisionally define a *self* as an agent plus his facticity, taking this to be analogous to the theoretical content plus initial conditions. A self is revealed by an experimenter's ability to exploit its

---

[2] This is a little crude, and finer distinctions can be made within what I have called heteronomous cooperation. For instance, there seems to be a difference between the members of the road gang, that do not have to be postulated as having intentional states at all (although, of course, they do have such states), members of the battalion who are allowed limited intentional states, at least to the extent that they can be said to be following rules, and the traditional manipulation cases involving neurophysiologists with their arrays of probes and electrodes. As you might expect, the 'agents' are not autonomous in heteronomous cooperation and are barely recognizable as agents at all. It is only with more advanced forms of cooperation that the question of autonomy becomes intelligible.

causal powers to produce predictable results. Hence, collective action is an experiment. Take the battalion example. The commander performs an experiment by manipulating the variables that he believes are causally relevant to the actions of his theoretical entities: he issues orders to his companies. They march as expected. Thus, he is warranted in his belief that a certain phenomena, namely a march, can be reproduced by these manipulations, and that his ontology of battalions and companies is sound.

At this stage there is no reason for this ontology to include human beings, intentional subjects, or anything more microscopic than a company. Obviously, this does not mean that there aren't any, but only that they are redundant to the account of what happened. Suppose, however, that one such company does not behave as expected. Then, the commander must look for hidden variables. These variables may be hidden at the micro-level, at which point the commander may be forced to start seeing the constituent elements of his company, namely his soldiers. But, there is still no need at this stage to allow for independent behavior on the part of his soldiers. Rather, he has a kind of operational definition of his men according to their functional role.

This leads to a familiar critique of operationalism that, when some particular experimental apparatus t is taken to warrant us in positing an electron, this is a mistake, and that all we should really posit is an electron-ish entity as revealed in apparatus t, and name it a t-electron. The self that is revealed in an experiment should not be taken as a 'real', 'true' or 'core' self, but rather as a self characterized entirely by its functional role, as a soldier-self or a waiter-self. I will call this a t-self, and claim that these should be taken to be real in a robust sense when its causal powers can be used in the design of experiments. Obviously, what goes for a self, also goes for a group, e.g., a company[3].

---

[3] A t-self is composed of action-sentences requiring explaining plus some theoretical terms and axioms contextually introduced to explain it and, more importantly, to use it on other relevantly similar occasions, for instance in the design of other experiments. What I am proposing here is so-called *entity realism* with regard to t-selfs and groups. Neither a self nor a t-self are quite the same thing as an agent. One way of noting the difference is that agents are usually taken to possess a much wider range of intentional properties, whereas t-

The anomalous behavior may not be due to variables hidden in the microstructure but to complex macro-level interactions, like feedback effects. Here the heteronomous approach of the commander reaches a natural limit. Issuing orders is no longer so effective where the variables can not be held constant, so the experiment must in some sense be allowed to become self-regulating, which is to say that some initiative must be given to the soldiers so that they can adapt. This means that at least some part of the march must be desired intrinsically by the soldiers, and that they cannot desire simply to follow orders. As an ideal, the soldiers will share the goal of its commander, so as to better coordinate their efforts to achieve it. Borrowing Dennett's terminology, we can say that the commander has to take an *intentional stance* towards his men. Even though other stances are possible we can still say that it is at this level, the level of the individual soldiers, and in more advanced forms of cooperation, that intentional properties start to be ascribed. This expands the theoretical content of the t-self, e.g., they must possess more complex dispositions than before, but does not require a change in our ontology.

This leads us to *intentional cooperation* and it is with this that we are mainly concerned. This will also be shown to involve 'thinner' and 'thicker' conceptions. What we are moving towards is a conception 'thick' enough to raise the question: does a group satisfying certain conditions constitute an agent in its own right?

There is a certain amount of common-sense to say that it does, purely from facts about our ordinary linguistic usage. It is common to attribute some action or attitude to a collective, e.g., "England defeats Germany on penalties", "Russell and Whitehead wrote the *Principia Mathematica*". The question becomes whether all such attributions can be explained away as figures of speech such that it is only in a metaphorical sense that a collective can be considered as an agent, rather as it is by

---

selfs are only taken to possess the properties necessary for the collective action under the experimental conditions given by t. Thus, as far as cooperation is concerned, what we are dealing with is t-selfs.

stipulation that some institutions are classified as 'legal persons'. What kind of things must we thus be able to explain away?

Firstly, we must be able to explain how we can attribute something to a collective without attributing it to *all* of its members. It was not everybody who qualifies as a British national that defeated everybody who qualifies as a German national on penalties, but only some, in this case the respective football teams. Secondly, it might seem false to attribute something to *any* of its members. To say "Russell and Whitehead wrote the *Principia Mathematica*" does not seem equivalent to either a conjunction or a disjunction of the statements "Russell wrote the *Principia Mathematica*" and "Whitehead wrote the *Principia Mathematica*". Velleman (1997, 29-30), from whom I take this example, notes that such attempts to explain away collective attributions using only individualist concepts fail to incorporate features of groups such as, when a group is asked to make a decision about something, it is being asked to make a decision *as a group*. Thirdly, attitudes of a group can be radically discontinuous from the attitudes of its members, and it is possible that two groups with the same membership will make opposing decisions *as groups*.

We have, then, both individual and collective concepts of agents, intentions, and actions and it is our job to inquire as to the relation between these concepts. I begin by taking this relation to be *mereological supervenience*. The next question is whether this relation performs a complete reduction of the collective to the individual. Theories that argue for such a reduction are classified under *methodological individualism*.

The mereological supervenience relation, as it applies to action, should, I think, be taken as a relation between *act-trees*. An *act-tree* consists of a basic *act-token* and other act-tokens *level-generated*[4] by that act-token.

---

[4] *Act-trees, act-tokens*, and *level-generation* are technical terms introduced in Goldman (1970). Without going into details, *level-generation* is meant, in part, as an elucidation of the relation given in the by-locution "I X-ed by Y-ing". X and Y are *act-tokens* on the same *act-tree*. All act-tokens on the same act-tree are performed by same agent at the same time. For Goldman (and myself) each act-token is a distinct event. However, we can continue to use this terminology even if we prefer the Davidsonian view that these act-tokens are just different descriptions of the same event, in which case it is act-trees that are distinct events.

Everything on an act-tree is an act-token and is something that the agent does rather than 'suffers' or 'undergoes', but only the basic act-token is guaranteed to have been performed intentionally. The act-tree of the collective supervenes on the act-tree of the individual, but not necessarily in virtue of an act-token that the individual performs intentionally; a complete overlap of intentions between the macro- and micro-levels is not needed. For instance, a member of a battalion has been trained so that when he hears the officer shout a certain word he performs a certain sequence of bodily movements. These movements may be a mereological part of a larger action, but the soldier does not know *what* action; the content of his intention is only to follow the orders of his officer.

This concludes my account of the relation between collective actions and individual actions: the former mereologically supervene on the latter. Most act-types need only a very weak sense of cooperation in order to be truly attributed to the collective. Other act-types, like reaching an agreement, do conceptually seem to require intentionality on the members' part. For either act-type *intentional cooperation* will be seen to possess features that weaker forms do not. A first step towards intentional cooperation is a *shared intention* or a *shared goal*: the group intends to X. The question is: how does the group-intention relate to the intentions of the group members? We provide the same answer as before: by mereological supervenience.

How are collective intentions related to individual intentions?

First of all, let us look at some attempts to build up collective intentionality from purely individualistic, or in Tuomela's terminology, *I-mode* resources. If this *methodological individualism* can succeed, then it seems that the principle of parsimony demands that we should reject any thoughts of irreducibly social facts.

The next footnote gives an example of such a translation. It should be borne in mind, though, that terms referring to events will be co-extensive on Davidson's theory that are not on Goldman's theory.

Bratman's account is of this type. He gives three roles to be played by what he terms *shared intentions:* coordinating activities, coordinating plans, and "a background framework that structures relevant bargaining" (Bratman 1993, 99). Intentions work by constraining plans to meet certain rationality conditions, e.g., means-end consistency, and are attitudes had by the individual agent to individual actions. Shared intentions have to be built out of these.

The first strategy Bratman considers is that the joint action is something that I want but do not strictly speaking intend. This seems connected to the idea that one can intend only one's own actions. Tuomela attributes such a view to Miller, who claims that although all the members of a collective may have an aim, and that this aim is the content of a conative attitude, we cannot *intend* such an aim strictly speaking. Tuomela finds this unjustified and says that this is an intention with a different kind of content, an *aim-intention* as opposed to an *action-intention*. The difference that Miller has identified but misconstrued is that *action-intentions*, i.e., an intention to raise my arm, can only be satisfied by the agent with the intention, in this case by me when I raise my arm. If someone else raises my arm then I did not satisfy my intention since I did not raise my arm intentionally. In the case of the *aim-intention*, when the aim is satisfied then it is satisfied for every member of the group who makes some kind of contribution even if this is only the mental act of *accepting* the aim [Tuomela n.d.(c), 23-24], and not just the members whose actions provide the finishing touches, so to speak, in bringing the aim to fruition. Bratman's strategy seems similar, which is to change the content of the intention from an *intention-to* X into an *intention-that* X. Such an intention can play the functional roles that he has previously specified and is thus coherent with his planning conception of agency. As a first approximation, then, our shared intention to J is composed of my intention that we J and your intention that we J, where J is a joint act-type (Bratman 1993, 101-102).

This condition is too weak, so Bratman (1993, 103-104) adds a common knowledge condition such that the intention is only shared if each agent knows that the other has the same intention and an efficacy condition such that this knowledge of the others' intentions is at least part of one's

reason for acting jointly. This efficacy is embodied in sub-plans that must match up to a point but not completely. They must 'mesh', and it is part of our shared intention that each others intentions be efficacious, or more specifically, it is built into the content of the intention to J such that it is an intention to J through meshing sub-plans (Bratman 1993, 105-106). Sub-plans mesh "just in case there is some way we could J that would not violate either of our sub-plans but would, rather, involve the successful execution of those sub-plans" (Bratman 1992, 32).

We can bargain about the best way to fill in those sub-plans. He says: "Each is rationally committed to pursuing means, and eschewing obstacles, to the complex goal of their J-ing by way of the other agent's relevant intention. Each aims at the efficacy of the intention of the other" (Bratman 1993, 109). Such rational commitment provides the framework for relevant bargaining, where perhaps one agent will have to help out the other in order to achieve the aim, if they can do so without undermining their own intention. He refers to these features as commitment to joint activity and commitment to mutual support (Bratman 1992, 328).

Where Bratman is concerned with shared intentions, Kutz is concerned with *shared goals*. Collective action involves what he calls *participatory intentions* as a common feature. These intentions are said to be *strategically responsive* when they are sensitive to what the agent thinks other agents are going to do. Joint actions involve such intentions, a shared goal, and usually but not always *mutual openness.*

Mutual openness is stronger than mutual belief and implies that we are favorably disposed to be responsive to what we think the other agent is going to do, so that I can adjust my plans.[5] In addition, both agents must conceive of their actions as contributing to the collective action, and the collective action must be due to the decision of each (Kutz 2000, 4-7).

---

[5] If, for instance, I want to fly to Bangkok jointly with my wife, and I know that she is planning to take such and such a flight, then I make my plans so that I take the same flight. Not only is the flight she is taking common knowledge between us, but she wants it to be common knowledge between us and wants me to be responsive to it. This seems much the same as Bratman's meshing sub-plans.

Although Kutz concedes that mutual expectation and responsiveness often exist in jointly intentional action, he claims that we should not insist on it. This makes the cooperation captured by Kutz's analysis weaker than Bratman's, as is Kutz's aim and for which reason he calls his analysis 'minimal'. For such a minimally but still jointly intentional action consider the joint action of saving a picnic. It starts to rain, and quite spontaneously one person grabs the food and the other grabs the crockery and cutlery:

> So joint action as such requires neither positive belief about others' intentions nor dispositions of responsiveness, since we can conceive of genuinely joint, if simple, forms of collective action in their absence so long as agents nonetheless act with participatory intentions. Only one further general condition seems to be required as part of the very concept of joint action: a condition of extensional overlap. It must be the same joint enterprise in which agents intentionally participate. . . .
>
> . . . Agents' intentions overlap – they *share goals* – when the collective end component of their participatory intentions refers to the same activity or outcome and when there is a non-empty intersection of the states of affairs satisfying those collective ends. (Kutz 2000, 20)

Shared goals move us a step closer to fully collective action, to *intentional cooperation*. Kutz gives an example. You are going to a friend's house for a quiet dinner and I am going there for a surprise party thrown for you. Our intentions overlap under the description of going to a friend's house, which satisfies both of our ends, so we have this as our shared goal. Just as an action can be intentional under one description and not under another, so also can it be jointly intentional under one description and not under another (Kutz 2000, 21).[6] The fact that the intentions only need to

[6] Kutz seems to prefer a coarse-grained way of talking where "I X-ed by Y-ing" just gives different descriptions of the act rather than distinct act-tokens. We can rephrase his idea using Goldman's terminology. We both have an act-tree: your intrinsic action-want is to go to my friend's house for a quiet meal whereas my intrinsic action-want is to go there for a surprise party thrown for you. The act-token of going to our friend's house is on both of our act-trees. Considering now the act-tree of the mereological sum consisting of the actions of everyone involved with the surprise party: this supervenes on my act-tree in virtue of my

overlap *extensionally* rather than *intensionally* means that this is still a very weak condition. If the intentions overlap *intensionally* as well, then this is clearly even more cooperative. I will call the version requiring only extensional overlap *minimally intentional cooperation*, and the version requiring intensional overlap *weakly intentional cooperation*.

A *participatory intention* has an individual role and a collective end. The relation of the individual act to the collective end might be expressive of one's membership of a collective, like wearing a business suit, or normative, complying with standards within the collective. It is the agent's conception of this relation that makes her intentions participatory – it must be seen as instrumental to the collective action and as generating some form of obligation through either formal rules or social expectations (Kutz 2000, 10-13); it leads to a normative commitment.

Bratman's view does not lead to a normative commitment. Who is correct here? Gilbert expresses the view that such a commitment must have a normative aspect such that if an agent breaks the intention then he should be rebuked. Bratman concedes that such a person is being unreasonable, but not that he is breaking an obligation. Often promises to each other may turn it into an obligation, but intentions can be shared without any promises being made. When promises *are* made, then we can say that the agents are normatively as well as rationally committed, and hence have additional motives to act jointly, but this is a level beyond mere shared intention (Bratman 1993, 110-112).

I am none too sure whether Bratman is right here or whether he has understood Gilbert's position. Remember that Bratman calls for *relevant bargaining*. Now, it seems to me implausible for agents to bargain with each other every time that one needs help to perform the joint action. There will be an expectation that some help will be offered automatically, simply as a

---

act-token of going to our friend's house and on your act-tree in virtue of your act-token of going to our friend's house. The collective act that can be called "attending a surprise party" can be in any participating person's action-plan apart from yours, because if attending a surprise party is a goal that you share, then it is logically impossible for it to be a surprise. Here we have a collective act that can only be *minimally intentional*.

result of the agents' commitment to the group. When bargaining is involved, sometimes a bargain will be struck, and sometimes it won't. In each of these three cases, the agent appealed to for help must judge the extent of his responsibilities, and this seems to me to be a normative judgment. If the person needs help because of bad preparation or negligence on their part, then the person who can help has to reevaluate the aim and decide whether it is worth the trouble, or whether to abandon the aim, possibly taking sanctions against the person who did not play their part in the joint action. In other words, reasoning has to go on at the evaluative as well as at the instrumental level. For this reason, I agree with Gilbert and prefer Kutz's theory to Bratman's.

Also, Bratman (1993, 111) seems to see Gilbert's approach as based on promises, but Gilbert explicitly rejects this view. On her view, the obligations operative are derived from *joint decisions.* As she points out (Gilbert 1983, 689), if they were promises, then they would not depend on one another. If I promise to walk the dog, and you promise to feed him, and you fail to carry out your promise, then it is not implied that I am released from my promise; my obligation to walk the dog persists. This is not the case with the kind of agreements involved in shared intentions. There, if one of the agents changes his mind, then the agreement is void. Kutz's view is also premised on decisions and seems to me the correct view.

This feature of interdependent commitment must consequently be achieved in a different way to promising, and the use of 'we' in attributing actions or attitudes to the group suggests this interdependence. Gilbert calls such a 'we' a *plural subject*. To enter into such a commitment, all parties must show their willingness to enter it under conditions of common knowledge (Gilbert 1983, 691-92). The obligations thereby derived are 'persisting' and can only disappear if the agreement is rescinded (Gilbert 1983, 700).

Gilbert's theory is a further move away from methodological individualism, but according to Velleman, the plural subjects formed by coordinated conditional commitments give us only a coordinated will, rather than a single will. Velleman wishes to show how a single will can be formed within Searle's framework. Searle, like Bratman, is an individualist, but

Velleman objects that this approach only works where a single agent has the authority to decide for the collective, but in this case the intention is not really shared; an intention settles something, and I cannot settle it if I do not have a say. The problem is to show how *each* agent can settle a *single* issue. I can settle part of the issue, lifting my end of the sofa and hoping that you lift yours, and together producing a single result. However, Velleman (1997, 29-35) is aiming for a more literal and stronger sense in which an intention can be shared.

According to Velleman, the individualistic response accounts for shared goals rather than shared intentions. By defining conditions under which we share the goal of lifting the sofa, and certain structural and functional relations between us, it hopes to reduce all collectivity into individuals. Velleman instead puts goals to one side and asks: what is an intention? He recounts, and endorses, Searle's answer that an intention is "a representation that caused action by representing itself as causing it" (Velleman 1997, 38), further remarking that this need not be a *mental* representation. Such a representation could be oral or written, perhaps formed by speech acts of all the participants, through which act they 'share' the intention. Gilbert's commitments may qualify as such a representation.

Velleman argues for this concept of an intention as follows. He has to show that something like a speech act satisfies the functional description that Searle gives it, which is to say that it causes what it represents itself as causing. To this end, he appeals to the motivation of understanding one's own action. Suppose that you are considering going for a walk, but you are sitting in your favorite chair, watching your favorite TV show, and it is raining outside. Yet you know that your best judgment is to go for a walk. How to combat this weakness of will? Velleman replies: by representing the fact that you are going for a walk, e.g., by telling your wife "I'm going for a walk". Rather than reveal inconsistency, you are motivated to actually go for a walk, and you do. This representation: a) causes you to go for a walk, and; b) represents itself as doing so. He admits that (b) requires more argument, and continues (Velleman 1997, 40):

> The agent who expresses an intention by saying "I'm going for a walk" does not represent the projected walk as something that was going to happen anyway, whether or not he had said so . . . [but] as something that is now going to happen precisely because of his hereby saying so. His statement thus differs from a report or prediction in that it doesn't purport to convey a truth independent of itself.

The next step is to note that a conditional commitment can be signaled in a representative (e.g., speech) act of the form "I will if you will". If this is responded to with the representative act "Then I will", then the antecedent of this conditional is true and their combination yields an unconditional commitment, which Velleman claims is a single token-representation. This token is literally the shared intention that we were aiming for. The conditional nature of the original commitment displays the fact that it is not settled only by the agent but gives the mechanism by which more than one agent can settle a single issue. Nor is it settled by two people saying "I am going for a walk", which only communicate independent intentions and not a shared intention, coordinated wills in contrast to a single will.

Ingenious as I find Velleman's account, I think it is mistaken. Although one might sometimes motivate or cause oneself to do something by saying it out loud to someone you expect to hold you to your word, it seems to me implausible that this is a common occurrence, so at best Velleman's account applies to a very limited number of cases, much more limited than what we would normally call a shared intention.

But I am not sure that it even applies to these. Although one may cause oneself to do something in the way he describes, the causal link is not direct enough. Searle's concept of causal self-referentiality is in part an account of a *proximate* cause. Here, the causal work done by the intentional content of the utterance "I am going for a walk" does not act proximately but acts *distally* through an intention to act consistently with how you have said you will act. It is this latter intention, under which the act of going for a walk is specified, that represents itself as causing the behavior that it causes, and this intention is mental.

The supervenience approach does not seem to require the single will that Velleman so enthusiastically pursues. Tuomela [n.d.(a), 24-26] describes the supervenience of the group on its members as having to satisfy two claims. The first is:

1. The *embodiment* claim that whenever a group has an attitude or does something, there must be an attitude really in the supervenience base. A group must have *operative* members who make decisions and act on behalf of the group.

It must not be held that exactly the same attitude exists at the group-level as (somewhere) at the base-level, or that for any particular proposition, the group's belief in the proposition requires at least some members' belief in that same proposition. In fact, attitudes at the group-level can be radically discontinuous from those at the base-level.[7] I would further hazard the opinion that the attitudes in the supervenience base do not have to be doxastic. This is consistent with Tuomela's distinction between belief and *acceptance*: the latter is an action and performative whereas the former is experiential and dispositional. When we attribute a belief to a group, we are saying that the members accept it as the view of the group, and not that the members believe it is true; scenarios can be constructed in which no members at all believe it to be true. This is also true for what Tuomela calls *positional beliefs*: beliefs that may be held to be false by those holding positions in the social hierarchy but are accepted by them for the sake of, and on behalf of, the group. Summing up Tuomela [n.d.(a), 10] says:

> A group is taken to believe something p if it accepts p as its view. This can only be the case if the group members or some of them, the operative ones, collectively or jointly accept p for the group. When they do so they must be acting correctly *qua* group members, viz. functioning in their positions in the group when the right social and normative circumstances obtain. The non-operative members must tacitly accept, or at least put up with, what the operative members accept as the group's views.

[7] This result has become well-known as Arrow's Paradox and is often used, e.g. by Pettit and Copp, as a support in arguments for various kinds of autonomy of collectives.

They need not even have detailed knowledge about what is being so accepted.

Since acceptance is an act, we can ask whether such an act was intentional. If it is an intentional joint action, then we can speak of joint acceptance, and when a proposition is jointly accepted then such beliefs qualify as *mutual beliefs* and *group-binding*. Group-binding beliefs are normative [Tuomela n.d.(a), 11].

The second claim is [Tuomela n.d.(a), 24-26]:

2. The *determination* claim that such attitudes on the part of the members determine in a non-causal sense that of the group.

Armed with his notion of joint acceptance, Tuomela posits what he calls a *jointness* level between the individual and group level, consisting not only of mutual beliefs but of other attitudes formed in the same way. For example, group-intentions are joint acceptances of optative propositions. Group-level properties are based on collective acceptance, through which properties in the base and jointness level are 'conventionally', rather than causally, connected; in other words, collective acceptance is the means by which the *determination* claim is satisfied, the means by which reason is collectivized in a manner constitutive of collectively rational action.

Joint acceptance must also satisfy two requirements. The first is:

a) The authority requirement states that the right social and normative circumstances, as given by formal and informal rules, must be present.

These rules define what the position-holder has the authority to do in virtue of the position he holds. If the position-holder exceeds his authority, then this does not count and any propositions he accepts in this mode are not binding on the group; the group has not acted intentionally (Tuomela 1989, 480-81). For instance, if a minor functionary of an organization signs a contract on 'behalf' of the organization, then such a contract is not legally binding if, in so signing, the functionary broke the rules delimiting his authority. Underpinning these rules are "general constitutive rules concerning the purposes and functions of the collective" and "proper social

norms specifying his social roles" [Tuomela n.d.(a), 8]. This tells you how reason is collectivized.[8] It should be noted that when we have identified the way in which the authority requirement is satisfied, we have identified what I earlier called the t-self – the *position* of the person in question, and the rules determining what they can and cannot do, together constitute an operational definition of the person. The mereological sum on which the collective intentions supervene are not intentions of fully autonomous agents but of t-selfs, of theoretical entities. It is these that we can say exist when we successfully use them in the design of experiments, that is to say, when a group successfully acts collectively.

The second requirement is:

b) The intentionality requirement states that some act-types, of which collective acceptance is one, are joint act-types, and some higher animals just are capable of joint actions in some primitive sense not requiring concepts or language, e.g., lions hunting together. This is a disposition that has evolved in social creatures [Tuomela n.d.(b), 5].

This seems to be an attempt to evade a charge of circularity that might seem to obtain if jointly intentional action is defined in terms of joint acceptance and joint acceptance in terms of jointly intentional action, a joint act-token being "an action performed by several agents who suitably relate their individual actions to each other's actions in pursuit of some joint goal or in adherence to some common rules, practices, or the like" (Tuomela 1989, 472).[9] I think that the intentionality requirement is a mistake, and that

---

[8] Arrow's Paradox teaches us that collectivizing rationality so that a group's decisions are fully rational, in the sense of being deductively closed, is a non-trivial task, and Pettit (2003) thinks that groups satisfying this constraint can justifiably be thought of as autonomous rational agents. I deny this conclusion and hold that groups must hold intentional properties, and not just some functional equivalent of an intentional property, in its own right before we should make this conclusion. In "The Weak Collective Agential Autonomy Thesis" I argued that as long as a distributive analysis of the attitude and action ascription is possible, we should stay with methodological individualism.

[9] The circularity objection applies to both Kutz and Tuomela, since they both take some act-types to be irreducibly collective, one of their reasons being that it is only by conceiving of oneself as part of a collective and as contributing to collective actions that problems in decision theory like the Prisoner's Dilemma can be solved, and that people do in fact take the group view when facing such dilemmas (Tuomela 2004, 7). I think that this only shows

we should not need biologically primitive and irreducibly collective forms of intentionality, either in the act-type, as suggested by Kutz and Tuomela, or in the form of the attitude, as suggested by Searle. For one thing, although this may not be an important point, it implies, since joint acceptances are taken to imply normative commitments, that higher animals and not just human beings have normative commitments. More importantly, we can account for the same facts when the only difference between the stronger kind of cooperation and the weaker forms is the type of supervenience relation. Where there is *person-wise* supervenience, i.e., the group's intentions and acts are some simple function of each member's intentions and acts, then attributions to the group can be distributed unproblematically to each member; hence, "The choir sat down" can be paraphrased as "Peter sat down", "Paul sat down", and so on for each member of the choir. Where there is *clique-wise* supervenience, we have to take into account the factors named in the authority requirement in a more complex way, so "America declared war" cannot be paraphrased into "Peter declared war", "Paul declared war" etc. because neither Peter nor Paul have the authority to declare war, but their intentions and acts can still be parts – as long as they are not indifferent to the outcome and support the action at least mentally – of the mereological sum on which the group's intentions and acts supervene. Some paraphrase into singular attributions and factual statements is still possible even when not all of the members of the collective participate actively in the action. For instance, we might say "Peter accepts r as a reason for declaring war", "Paul accepts r as a reason for declaring war" etc., corresponding to Tuomela's notion of collective

---

that we conceive of ourselves as part of a collective, without having to add that the collective, or its actions and attitudes, are irreducible.

Another possible source of the felt need for irreducibly collective act-types is perhaps the idea that verbs like 'to surround' seem to have as part of their concept the fact that they must be performed by a plurality. But I think that this is just a contingent fact about human beings, and that if we existed as soap bubbles or as pure energy then we could quite intelligibly speak of ourselves as surrounding. My conclusion is that we can do without such special act-types.

acceptance. Exactly what this paraphrase will be depends on the constitution of the group.[10]

How do Kutz's and Tuomela's views accord with the questions posed at the start? Are groups an irreducible part of the ontological furniture? They both posit irreducible act-types, but despite this it can be objected that they are too reductive. It is irreducible in terms of content, but not in terms of form. Searle gives the example of some business students who all believe that the way to benefit humanity is through rational self-interest and act appropriately, expecting the other students to act the same way. Despite intentionality and mutual belief, we would not say that they are acting jointly. However, if they made a pact to help humanity by acting in the same way, then they *would* be acting jointly. According to Searle, this can only be accounted for by a difference in form. Kutz responds that this puts the difference in the wrong place, and that the difference is between intending something and simply knowing that it will come about (Kutz 2000, 15-16). In Tuomela and Goldman's terms, they do not accept the optative proposition "We will benefit humanity", or have what Tuomela (1989, 486-87) calls a *conduct-plan*. Nor, Kutz seems to say, do they even accept the individualistic "I will benefit humanity", but only see it as a side-effect of their intention to help themselves.

By making this move, it seems to me that Kutz has simply changed the question. Let us suppose, as Searle probably intended, that helping humanity is their reason for acting as seems to me implied when Kutz stipulates that "each student believes that each student believes this [the doctrine of rational self-interest] to be true and will act upon it" (Kutz 2000, 15) and that they would individually believe "I will benefit humanity." The question now is: do the students as a group jointly intend to benefit humanity? Searle says no, and argues that no analysis in terms of mutual beliefs will be able to account for this difference. With this I am inclined to agree; we need more than beliefs. However, the analyses of Tuomela (and

---

[10] The *jointness* level, if we wish to preserve this idea, is simply the event-trees that supervene on the act-trees and attitudes of members of the cliques.

also of Bratman) allow for conative as well as doxastic elements, and these, I think, can do the job required.

How are collective agents related to individual agents?

Tuomela allows for collectives without metaphysical extravagance, noting that, since 'we' is occurring in an intensional context, we do not have to suppose that it really exists but has an "intentional inexistence" (Tuomela 2004, 22):

> We-mode talk requires having joint attitudes as a group and acting as a group. Thus the concept of group is referred to and relied upon here. However, this does not entail that groups must exist as entities. It suffices that they have "intentional inexistence" (as Brentano put it), viz. occur in intentional mental contents and thus lead people to act in relevant ways . . . What is ontically required to exist is the "jointness level", viz., joint actions, joint intentions, joint preferences, mutual beliefs, and other joint attitudes must be taken to exist. Group members' functioning in the we-mode requires that we (and they) attribute goals and standards (etc.) to groups, although they in my account do not literally, in an ontic sense have them.
>
> This means that I can have a we-intention even where "we" has no referent, consistently with the internalist constraint that intentionality is the way it is irrespective of the way things are in the world and would be the same even in the solipsistic scenario where there are no other agents.

I have argued to the contrary that the "we" is reducible to singular attributions, which clearly do have referents, but what they refer to are theoretical entities. The fact that we can successfully manipulate them entitles us to treat the semantics of such attributions in a realistic way. However, it should be noted that this means that "we" is always parasitic on "they"; the person making the attribution is always taking the point of view of the experimenter outside of the experiment itself, even when they (or, more properly, their t-selfs) are themselves one of entities manipulated in the experiment.

Collective attributions like "We believe that we ourselves wrote *Principia Mathematica*" and "We wrote *Principia Mathematica*" are analyzed in the following way. Since we hold the embodiment claim, there is some operative member to whom we can ascribe some property relevant to the writing of *Principia Mathematica*. I hold also that ascription of a property to an individual is also to ascribe a property (not necessarily the same property) to any mereological sum or clique of which it is a member, and conversely that ascription of a property to a collective *ipso facto* ascribes properties to all of its members; the mereological supervenience relation will tell you exactly how to distribute these ascriptions. This applies in so far as one's ascription to the operative member takes that person, or more properly, the t-self, as an intentional object, i.e., it is from a third-person point of view, even if I myself am the operative member. [11]

Thus, what we have is not a *we-intention* but a *they-intention*, and in order to get to "We believe that we ourselves wrote *Principia Mathematica*" or "We wrote *Principia Mathematica*" from its third-person cognate, a further belief[12] to the effect that I am a member of this collective is necessary. I think that this is the 'thickest' concept of cooperation we can get.

---

[11] I hold that there are certain kinds of indeterminacy in the contents of our intentional states, such that ascriptions to individuals also imply, in some sense that does not presuppose that both ascriptions are true, ascriptions to collectives, and ascriptions to collectives imply in the same sense ascriptions to individuals. I cannot argue for this here, but note that it is approximately the same as running the embodiment claim in both directions – if a member of a collective does something, then the collective does something, and if a collective does something, all of its members do something, although we may not be able to say what any of these act-types are except by stipulation.
The agent can take a dual role, as the experimenter and as one of the entities experimented on. Each theory of the t-self is something exterior to it, postulated by the agent or by the other members of the collective, each regarding themselves as experimenters. Membership of the group is analytically prior to the individual's sense of agency; hence, the agent is always partly alienated from his t-selfs. If an agent theorizes over his t-selfs, this is in order to justify his actions to other people who would hold him accountable for them, and he may well have different theories depending on who those people are asking so insistently "Why did you do that?"

[12] I think this has to be a *de se* or *self-locating* belief.

Conclusion

I endorse my variant of Tuomela's theory for *strongly intentional cooperation* and Kutz's theory for *minimally* and *weakly intentional cooperation.* However, I think that a problem has emerged. We started off with *heteronomous cooperation* and, by advancing into higher and higher levels of cooperation, assumed that our autonomy would increase. However, *strongly intentional cooperation* has brought in further obstructions to our autonomy. It forces us to take a stance of detachment towards ourselves; our individual perspective has been taken over by the group perspective and forced us to see ourselves fundamentally as part of the group and bound by its norms, forced by our positions in the group to take as reasons for action propositions that we do not believe are reasons for actions, or even in some cases to be true. Putting collective autonomy to one side, it becomes a real issue whether there is such a thing as autonomy for the individual.

It might seem that this is not even an intelligible question. It seems like it might be asked "What is an autonomous X?" where X may be a soldier or a waiter, or a planning committee or a football team, in which cases the answer is "A soldier whose behavior requires the postulation of a certain amount and kind of independence in order to explain it" etc., but it is not clear that it is even intelligible to ask whether there is such a thing as being autonomous *simpliciter*, divorced from any sortal term or *position*. I think that if there is such a thing, it can only be found through something like existentialism.[13]

---

[13] Critics would probably say that this sortal autonomy is the only kind we need. Most of the discussions of the moral autonomy of collectives, for instance, do not seem to require more than this. I am tempted to think that autonomy of the more problematic kind is connected to responsibility and responsibility to giving reasons and reactive attitudes, where giving reasons are collective actions and reactive attitudes are we-attitudes (or they-attitudes!). This seems to lead us full circle. I think that this reaches a kind of practical terminus in situations of *minimally intentional cooperation* where only one person, the one who 'designs the experiment' and decides where the 'extensional overlaps' occur, can be autonomous. Thus, my autonomy over strongly collective actions depends on my autonomy over the justificatory practices concerning that action, which practices must degenerate into a kind of collective action where only I have the autonomy. I think that this strategy will also halt the regress mentioned in the next paragraph and accounts for the fact that much of

Some philosophers would probably say at this point that an agent's autonomy is not compromised if they have accepted the terms that define their membership of the group and have the option of leaving the group. Under Islamic law a man can divorce his wife simply by saying "I divorce you", thus leaving one particular group consisting of himself and his wife. But how did that law come about? Neither the man nor his wife had any choice in the matter. This is a problem quite generally: one might put in a procedure designed to bring about the fairest outcome, e.g., majority vote, but do you also have a procedure to decide whether or not to implement this procedure or another? Eventually you must reach a bedrock of what is just given. Tuomela acknowledges this, and I think this is partly behind Tuomela's intentionality requirement, that I think he borrows from Searle, that some intentionality, including some collective intentionality, is just a biologically primitive phenomenon. This manoeuvre seems to me an over-reaction.

Earlier I said that once properties of a certain kind have emerged at one level then ascriptions of the same kinds of properties at higher levels can be reduced to the lower level. Are there any such properties emerging at the collective level? If there were irreducible joint act-types, or irreducible we-attitudes, then we might have to abandon methodological individualism. But I hope to have shown that we do not need to accept these; we can continue to say that the higher level mereologically supervenes on the lower level, although some cases require *clique-wise* supervenience. A group *has* an action-plan in virtue of its members' joint acceptance of an action-plan and is not an agent in its own right for the same reason that we suspected all along, that is, because it is not an embodied mind.

References

Bratman, Michael. 1993. Shared intention. *Ethics* Vol. 104 No. 1

___1992. Shared cooperative activity. *The Philosophical Review*, Vol. 101 No. 2

---

our constitution must simply be taken as given without having to bring in *sui generis* forms of intentionality.

Copp, David. 2007. The collective moral autonomy thesis. *Journal of Social Philosophy* Vol. XXXVIII No. 3

Gilbert, Margaret. 1993. Agreements, coercion, and obligation. *Ethics* Vol. 103 No. 4

Goldman, Alvin I. 1970. *A theory of human action*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kutz, Christopher. 2000. Acting together. *Philosophy and Phenomenological Research* Vol. 61 No. 1

Pettit, Philip. 2003. Groups with minds of their own. [Available online]. www.princeton.edu/~ppettit/papers/GroupMinds.pdf. Accessed 3 August 2007

Tuomela, Raimo. 1989. Action by collectives. *Philosophical Perspectives* Vol. 3 Philosophy of Mind and Action Theory

___. n.d.(a). Collective intentionality and social agents. [Available online]. www.valt.helsinki.fi/kfil/matti/tuomela.pdf. Accessed 27 July 2007.

___. 2004. Cooperation and the we-perspective. [Available online]. www.valt.helsinki.fi/staff/tuomela/papers/St_Gallen_article.htm. Accessed 7 August 2007

___. n.d.(b). Joint action. [Available online]. www.valt.helsinki.fi/ staff/tuomela/papers/Joint_Action.htm. Accessed 7 August 2007

___.n.d.(c). We-intentions revisited. [Available online]. www.valt. helsinki.fi/staff/tuomela/papers/We-Intentions_Revisited.htm. Accessed 27 July 2007.

Velleman, J. David. 1997. How to share an intention. *Philosophy and Phenomenological Research* Vol. 57 No. 1

Von Mises, Richard. 1968. *Positivism*. New York: Dover Publications.

Intuition and synonymy – the extension of coverage of a concept

An analytical approach

Marcel BODEA*
Babes-Bolyai University Cluj-Napoca

Abstract:

This article has as objective a particular analysis, from the perspective of linguistic synonymy, of the report *common language / mathematical language*. The analysis is based on a case of study: "the extension of coverage of a concept". The case of study approached has mainly an algebraic content. The interpretation of the case of study also requires a semiotic frame. We introduced a "compliancy condition of the senses". The compliancy condition of the senses means, that the sense of the expressions in the two different languages: mathematic and common is given by the sense in the reference language, i.e. the common one.

Keywords: synonymy, intuition, linguistic symmetry, lexical equivalence, equivalence of meaning, intuitive notion the infinite ("∞"), finite/infinite sets, function, bijectivity (one-to-one correspondence), cardinality, intuitive paradoxes.

About *synonymy* – preliminary characterization

The analytical intentions of clarifying leave the field clear for analogies with the condition of explicitly formulating the boundaries of the analysis in each particular case. The case of study approached has mainly an algebraic content. The interpretation of the case of study also requires a semiotic frame. Through some precisions and examples from the text itself and from the notes, we have tried an indirect circumlocution of this frame especially at the formal and terminological level. The present analysis

* E-mail: bodeamarcel@hotmail.com.

begins with a characterization of synonymy from the point of view of the philosophical analysis of the language. The substitution through synonymy is on the one hand, from the linguistic point of view, a potentiality of expression and on the other hand, from the philosophic point of view, an abstract relation of the language. Synonyms are concrete contextual acts (of this potentiality) which preserve, locally: the object –referential significance of the words and globally: the meaning of the propositions. As reference we had in mind the linguistic approach of *synonymy* in the context of the spoken language and in the literary context. We will try the construction and respectively the application of the notion of *synonymy* in the philosophical analysis of the language, as a rigorous theoretical notion.

What are *synonyms*?

> […] *synonyms,* i.e. the lexical equivalence to express the same notion, […] there are often semantic nuances between synonyms, […]
>
> *Synonyms* are those words with almost identical meaning, which can be interchanged, which can alternate in a given context, without changing the global meaning of the message.
>
> […] a given fact is that everyone uses as equivalences words such as *work – labor*; *to go – to leave*; *to arrive – to get there*…;[1]

We will postulate the defining characteristics of the synonyms – making abstraction of estimations, nuances etc. which will be considered in the following text:

1) *Synonyms* are *words*. In a language, synonymy is a relation between words, but only in a propositional frame. (Synonyms are no propositions![2]). In order to be able to talk about a non trivial relation of synonymy, "the

[1] Bulgăr Gh., *Dicţionar de sinonime*; Editura Lucman, 2004, Bucureşti, pp. 5-6.

[2] This doesn't mean that an extension of the relation of synonymy for linguistic objects won't be possible. Generalization imply reconfigurations at the level of semantics of the propositions, of the interpropositional relations, of the meaning of phrases, etc. limiting synonyms is in compliance with the normal linguistic sense, and without being simplistic, this is philosophically useful through its consequences.

difference condition" is imposed as a requirement: synonyms are *different words.*[3]

2) *Synonyms* have as "semantic reference the same *object* [referent]. Through "semantic reference" we define here the univocal correspondence between an implicit word in a propositional frame (a proposition explicitly formulated or not) and a well defined object.[4] The referential correspondence, of semiotic nature, will be in this way a "local" one, limited here to the level of a propositional announcement (the relationship between a word in a proposition with a an object).[5] With the new requirement, the condition of synonymy becomes "stronger": in order to talk about *synonyms* we need to have *different words* wit*h the same objectival reference* (in other words, if two words are synonyms, then, they are lexically different and they have the same semantic reference [referent]). This is a required condition to make the synonymy possible. We introduce hereby a criterion, a "condition of difference" between words with the same reference: two words with the same objectival reference are different if there is one proposition where their substitution leads to different propositional meanings.[6]

---

[3] "The condition of difference" between words with the same semantic reference will be defined and exemplified in the text.

[4] "Objects" can be mainly of any nature, however general or identifiable. In other terminology, for these objects of reference we can use the word "referent". Emphasizing the meaning, "'object" has a more ontical meaning, whereas "referent" has a more semiotic meaning. In the analysis of this article, the nature of the objects is of more interest than their semiotic statute (presumed as implicit) and the syntagmas *object of reference, objectival reference, semantic reference, objectival significance.* Sometimes, in order to highlight also semiotic aspects in the context, the word *referent* will be put between brackets. The interest for the nature of objects to the disadvantage of the semiotic function doesn't imply the development of some "ontological engagements" in the text. To be noticed for this approach, that there are *words* without any *objectival reference* such as the word "and".

[5] The intentions of this study are not primordial semiotic. This is why we opted here for a simple "semantic correspondence" from a classic point of view: words are related to objects through a relatively univocal and well defined description.

[6] "Difference" between words is more than a difference of "signs" or of a form of symbolic representation (the way we spell or write words). Explanatory remarks about the differences of propositional meaning in the given conditions will be made in the text.

3) Between *synonyms* there is a "relation" of *lexical equivalence*. The substitution of some words in a proposition requires a condition of invariance, of "linguistic symmetry", in order to be a substitution through synonymy. *the linguistic substitution through synonymy in a proposition preserves the propositional meaning.*[7] In other words, substitution through synonymy means changing words in propositional form without changing the meaning of the expression.

In few words: The synonymy is the semantic relation that holds between two words that can (in a given context) express the same meaning; synonymy is a lexical relation that means sameness of meaning. Synonyms are similar, but not identical.

For more specificity, let us make the following formal comment, which emphasizes the *relation* (relating). Because the formal aspect emphasizes the signs (symbols), we will generally presume here, the possibility of signs to render something independent of them. The (explicit) function of representation of signs will be called in this case (their) significance.

In the current text, from the perspective of the objectives of the proposed analysis, we consider *reference* as a relation between two entities: a "linguistic entity", a "word" (usual acceptance) and a well defined "objectival entity", whose ontical nature can be however general or specific.[8] We will symbolize this relationship of reference as follows: (c,o). We will define in the following *a relationship between words* "with

---

[7] Bulgăr, 2004, p.5: "The segment of communication: 'His father has built a big house' can be reformulated as 'His father has built a large home', in order to see that each term from the first construction has a quite precise but, a different equivalent, in the second construction, still, the meaning of the idea hasn't changed." We gave this example for two reasons. First of all: because it highlights the preservation of the propositional sense (more exactly the preservation of the sense of an expression). Second of all: because a part of its expression is suggestive from the perspective of the case studies. Thus, "*each term from the first construction has a quite precise but, a different equivalent, in the second construction"* – is a usual wording in the common language, referring to a simple example of a relation of (lexical) equivalence between synonyms.

[8] For the simplicity of expression, "the objectival entity" will be also called in the following "objectival reference".

objectival reference (supposing that they are in relationship with "well defined objectival entities")".

Let c_1 and c_2 be two words. We will say about c_1 and c_2 that they are related (in a certain type of relation [for the time being expressed only algebraically]) if c_1 has an objectival reference (for example o_1) and c_2 also has an objectival reference (for example o_2). If the objectival reference is the same, then the relation is (algebraically) of equivalence. In this hypothesis"(c_1, c_2) means (c_1, o) and (c_2, o)". Let us show that the relation between words, redefined in this way, is an algebraic relation of equivalence.[9]

i. The relation is reflexive: indeed, $(c, c) \Leftrightarrow (c, o)$ and (c, o). In other words, a word has the same reference (this is the hypothesis in which a word doesn't change its reference).

ii. The relation is transitive: if (c_1, c_2) and (c_2, c_3), then we have the relations (c_1, o) and , respectively (c_2, o) and (c_3, o), from here we can write "(c_1, o) and (c_3, o)", which according to the definition of the relation between words, it formally means: (c_1, c_3). So (c_1, c_2) and $(c_2, c_3) \Rightarrow (c_1, c_3)$. In other words: if a word is in relation with a second word and this is in relation with a third word, then the first word and the third word are in relation. Transitivity means in this case "preserving the reference" through words.

iii. The relation is symmetrical: "(c_1, c_2) means: (c_1, o) and (c_2, o)" and since words don't change their reference, the order of expression of the words' reference is: "the word c_2 with the reference o" expressed before the "word c_1 with the reference o", doesn't modify their reference: (c_2, o) and (c_1, o) which is equivalent with (c_2, c_1). So $(c_1, c_2) \Rightarrow (c_2, c_1)$.

In this description, the algebraic relation of equivalence can be semantically interpreted through the "same referent". Philosophically we can make a convention of language by affirming that the algebraic equivalence is a "week equivalence" between words, sustained only at the

---

[9] The following demonstration is based on the way we normally talk, the way the words are usually used in the language, in other words, a linguistic pragmatism. A more detailed analysis shows how many tacit/unspoken assumptions are present in the regular speech.

"ontical" level. This "week equivalence" might be considered a *lexical equivalence* between words with the same referent. However, in the present analysis, synonymy will be considered at the level of a "stronger" lexical equivalence, implying, beside the necessary condition of the identity of the objectival reference, also the requirement of preserving the propositional meaning: equivalence of meaning.

A word is also assumed in a proposition, this is why we will simply state that it is in a propositional (linguistic) relation. We will symbolize this propositional relation as follows: $((c,o),p)$ and we will interpret it: the word *c* wit the reference *o* is in the proposition *p*. Above this, we usually say that a proposition has a sense. Nevertheless, we will consider, more precisely, that actually the complex relation $((c,o),p)$, which we will call "expression", has a sense. In this representation, *p* has a more general significance: it is the *form of expression*. In a language, *the form of expression* is given by the lexical repertoire and by the syntactic rules of the language. An expression symbolized in this way is a very general representation. However, in a certain language, this general linguistic relation has different specific representations. We introduce in the following a relation of meaning: $(((c,o),p),s)$, notation which we will interpret: an expression generated by a word *c* with the reference *o* integrated in a proposition *p* has the sense *s* associated. Through the sense of an expression it is meant here the sense of a form of expression. Moreover, for the simplicity of the language, we will use next the expression: the sense of the proposition *p* given by the word *c* for the "expression with sense $(((c,o),p),s)$". Let us have the following premise of characterizing the sense: "the sense of a proposition *p*, where the word *c* is being emphasized with an objectival reference *c*, is rendered by *the way in which the proposition exposes the reference."* For us to formally correlate the sense with the proposition and to underline the linguistic aspects, we will associate to the general relation of sense another formal representation: $(((c,o),p),s)\mapsto((c,o),p(c))$, where $p(c)$ is the sense of the proposition *p*, given by the explicit way throughout which

the sentence exposes the reference.[10] In these formal descriptions is present implicitly an assumption which we will explain: *there is o* if there is no reference object, there is no expression and we can't talk about the sense of the expression. In other words, in the present text, from the point of view of the propositional language used for expression, the propositions where a word has no well defined objectival reference [referent] have no meaning.[11] Moreover, those propositions and their derivate propositions are not actually expressions; these are not handled as propositions.[12] Remark (with the presumption of existence of the objectival reference): we accept that the way in which we use the language allows us sometimes to say naturally, without any preparatory theoretical considerations: "We speak about the same thing but in different situations" or "we speak about the same thing but from different perspectives", etc. in this article, this kind of a possibility formed by applying the daily language, stands - without any prior definitions – at the basis of characterizing the meaning of a proposition as *the way in which the proposition exposes the reference* (which means that we can distinguish between different ways of exposure without appealing to explicit criteria of differentiation.). In short, *the sense* is the meaning of an expression implied by its linguistic use in order to express the relation of a word with its reference. Furthermore, in the current text, because of the difficulty of some themes, such as that of the sense of expressions (propositions), in order to reduce ambiguities and to confer clarity and consistency to the central ideas, we will introduce in certain circumstances "compliancy conditions of the senses". "The compliancy conditions of the

[10] These formal descriptions will be exemplified in the text.

[11] It is actually implied here the perspective of the entire philosophical analysis.

[12] They might be interpreted as nonsense *propositions* in the Wittgenstein way of approaching the language proposed by Tractatus: *propositions with meaning, propositions with no meaning, nonsense propositions.* [Regarding these aspects we refer to Wittgenstein Ludwig, *Tractatus logico-philosophicus*; Transl. Dumitru M. & Flonta M., Editura Humanitas, Bucureşti, 1997, "În ajutorul cititorului " – M. Flonta, pp. 47-56.] The present analysis does not propose a parallel approach with the mentioned text. It introduces a series of premises useful for a certain formal orientation of conceptually clarifying the synonymy.

senses" for certain propositions of a language must be defined.[13] For example, in the proposed hypothesis of propositional sense, a possible "compliancy condition of the senses" is given by the exclusive interest for the objectival reference.

In order to emphasize certain features of the synonymy, we introduce at this point of the analysis the following proposition: "In a given proposition, a certain word *c* with the objectival reference *o* expresses this reference in a unique way"; consequently, the given proposition has a one and only one sense, it does not change its sense and it becomes obvious that a word is synonym with itself. This proposition implies however a certain separation from contexts, an 'approximation', in which the influence of the propositional contexts over the propositional sense is being disregarded; through this simplification, it is here implicitly supposed an unchanged given context, according to which the analysis will be made.[14]

In the same formal spirit, the relation of synonymy will be regarded as a relation of the type (c_i, c_j) determined by the specificity of the relations $(((c_i, o_i), p_i), s_i)$, respectively $(((c_j, o_j), p_j), s_j)$. A relation (c_i, c_j) is of synonymy if following two relations exist: $(((c_i, o), p_i), s)$ and $(((c_j, o), p_j), s)$. Hence, two words c_i and c_j are synonyms if they have the same reference *o* [referent] and the expressions generated through their substitution in a proposition *p* have the same sense *s*. The condition of preserving the sense is strongly related to the presupposition of existence of a language relation, which is explicitly demonstrated by structuring the language in (different) linguistic expressions. This is no more than an algebraic expression,

---

[13] These hypotheses are inspired, on the one hand from the way in which, within some languages (common or formal) different propositions (expressions) are considered to have the same meaning, and on the other hand, from the way propositions from different languages get to have, in report with a given or built reference language, the same meaning (for example the common language is often a reference language). To establish criteria, formal, logic or mathematical, as references for expressing the *consensus* of a proposition, is not an objective, not even a secondary one for the present article. However, somehow, the "consensus" needs to be affirmed or rejected.

[14] An example, in which the meaning of a proposition changes with the context, will be given along with the analysis.

simplified for the reference, sense and synonymy, which we consider suggestive.[15] Beside these features, to clarify the relation of synonymy, we will give explanatory examples, as different as possible from the synonymy. The examples are each time particular. Characterizing synonymy as a relation, requires however a certain degree of generality. This generality is however limited to a propositional frame and to an unchanged implicitly assumed propositional context. For clarification, we want to underline that synonyms are words which can be synonyms only in a proposition. We will convene that for a class of synonyms this generality is given by the semantic "bidimensionality": "vertical" is given by the report of the words (signs) with the *same* object [referent]; "horizontal" is given by the report of the words (signs) among themselves: the lexical equivalence (this lexical symmetry is directly associated with the preservation of the propositional sense).

We will highlight the existence in the mathematical and common languages of some similar "linguistic behaviors" and we will analyze several of their consequences. Is this similarity an argument in the affirmation of the unity of "factual thinking", expressed (here) through a common language and of the "symbolic thinking", expressed here through the formal (here mathematic) thinking?[16] Or is it perhaps just an argument in affirming the existence of some similar linguistic structures in the daily and mathematic expression (whose general significance might be exploited)?

In an interpretative way, the case of study follows a semiotic path, which was also synthesized by the philosopher David Clarke.he made a late

---

[15] As we can notice along this paper, this formal writing which simplifies the description, complicates in the same time, through its significance and interpretation, the philosophical –semantic analysis of the synonymy. Nevertheless, this is no drawback from the philosophical point of view, because the main philosophical objective is the conceptual clarification (here, through the algebraic analysis of the language.)

[16] The brackets imply that in the present analysis other forms of languages associated to the "factual thinking" are disregarded: e.g the language of the physics – of the "symbolic thinking" or the formal language of the smbolic logic

try (1987) of defining semiotics through the reduction at primitive signs, whose logical characteristics are rendered by the logic of la language.[17]

The present analysis also starts from several linguistic-philosophical assumptions. In these terms, the main assumption is the following: "*Synonymy* is a type of *relation* between words which refer to the same object [referent] and which preserve the sense of the proposition."[18] Is it possible and could it be also consistent - from the philosophical point of view - the extension of coverage of the synonymy relation beyond the internal limits of the common language, so that this would be applied to both the formal language and the relation between the languages? Following this direction we will introduce a double representation of the relation of synonymy: a relation of "internal synonymy" if the relation is between the words of the same language, for example in the common (or formal) language and respectively of "external synonymy" if the relation is between words of different languages, for example the common and the mathematical language. Referring to the relation of external synonymy, a question arises: could two terms from different languages have as semantic reference the same "object" [referent]? A word from the common language – in a proposition from the common language – has an object as semantic reference. Can a word from the mathematic language – in a proposition from the common language – have a semantic reference? And vice versa: for a given mathematic word, can a word from the common language have the same reference? Of course, the question about the possibility refers at least to some objects and words). Two examples illustrate in this text, contents and limits of both possibilities. 'the finitude" and "the equivalence classes". We notice that if this possibility exists, then the problem of *consensus* arises (consensus of agreement) between the according propositions from the two languages. The present analysis emphasizes the referential-objectival aspect.

---

[17] Deely John, *Bazele semioticii*; Transl. Mariana Neț, Editura All, Bucureşti, 1997, p. 4. We mention here, that for other semioticians, linguistics is not the only model for semiotics.
[18] From the general point of view, "the object" can become in turn a *relation.*

Let us accept at least for the beginning, the usage of the term "synonymy" in the following context. A simple but suggestive historic example of synonymy on the one hand intern, on the other hand extern is given by the "analytic geometry": the description in algebraic language of some geometric objects. To *describe* algebraically geometrical *objects* has proved to be, from the philosophic point of view, extremely profitable in mathematics, both at the level of the mathematic language, of "conceptual clarifications" and at the "ontical" level, of structure, of the relation number/space, etc. The expression "$y = ax + b$" is an algebraic term for the geometric term "line" (symbol *d*) and they both have as reference the same mathematic [geometric] object: the line, for which we also have an intuitive representation)[19] Within the current mathematic language we can talk about an internal synonymy, but at least historically, at the level of the distinction algebra-geometry, the synonymy is external, between languages considered different initially.

The distinction is important from the perspective of the consequences of the linguistic constructions through synonymy. In the present case of study, the emphasis is on the "external synonymies". The philosophic background of the present assumption is in its turn an assumption: "all the signs which refer to the same object [referent] can be put in relation among them."[20] The perspective of approaching the analysis in the case of study is that in which the "signs of the common language" and the signs of the "formal language meet and not the perspective in which they contrast. Consequently, we will talk in the text about mathematic objects in the common language. From a more general philosophical perspective, it is analyzed the possibility of speaking in a common language about formal and scientific objects, respectively, of using the common language as an important component of the specialized scientific language. The targeted purpose is in the end of analytical interest: conceptual clarifications. This

---

[19] Of course a and b are assumed to be fixed, so that we have the entire plan described by lines and the plan is another geometric object.

[20] We have the conviction that a philosophic analysis in this direction of the algebraic examples offered by *the internal* and resp. *external operations* can be philosophical profitable.

doesn't mean that the limits of the contact zone, here philosophic limits: intuitive and conceptual, are not looked for.

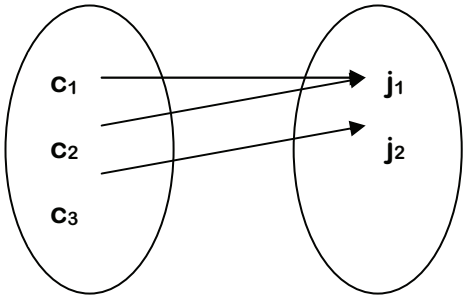
In this case of study: "the extension of coverage of a concept" through synonymy, starting from the common language, will be defined notions specific to the formal language of mathematics. In parallel with some expressions of the common language, the corresponding mathematic expressions will be formulated and some consequences will be analyzed. In the case of study presented, a relatively "strong" reduction of the *empiric* to finitude will be performed; the empiric being characterized *by* finitude.

The extension of coverage of a concept

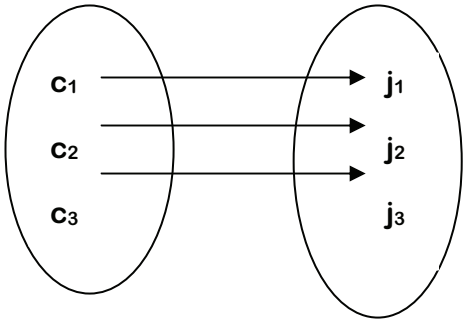
We will introduce at this point a first *compliancy condition of the senses* and we will say about two expressions from different languages that they have the same sense if there is the possibility of "translation", "articulation" and "representation" of expressions within the same reference language, in which the "converted" expressions will have the same sense. There is also the possibility for one of the languages to be the one of reference. For this case of study, the common language is the language of reference.

Let be the following description in the common language. There is a group of playing children and a corner with toys, where they are playing. The rule that, on the one hand all children should play and on the other hand that a child should play with a single toy is required. (Of course, a child cannot play with more than a toy, but two or more children can play with the same toy). Let us have the following drawings which render through images (somehow intuitively), the main playing reports of the children with toys.

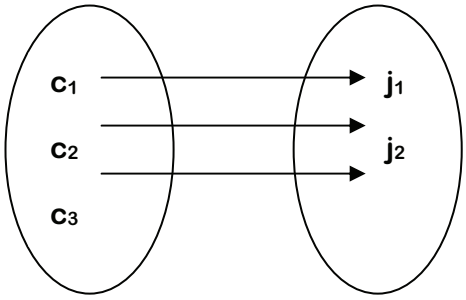
(D.1.)



(D.2.)



(D.3.)



Let us describe in the common language the factual situations represented by the above drawings.

(D.1.) There are three children and two toys. Two children play together with the same toy, one child plays alone. We notice that there are fewer toys than children.

(D.2.) There are three children and four toys. Each child plays alone with its toy. There is one toy left with which no child is playing. We notice that there are more toys than children.
(D.3.) There are three children and three toys. Each child is playing alone with its toy. We notice that there are as many toys as children.
The three described situations allow us to naturally affirm that: “there are as many toys as children if the number of toys is not smaller than the number of children and at the same time the number of toys is not bigger than the number of children and at the same time.

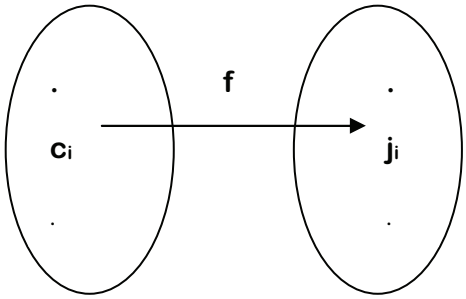
Looking comparatively at the drawings (D.1.) and (D.2.) respectively (D.3.) we will notice that, if there is the possibility for any two different children to play with different toys, then there are definitely no less toys than children, in other words the number of toys is no smaller that the number of children.

Looking comparatively at the drawings (D.2.) and (D.1) respectively (D.3.), we notice that, if there is the possibility at one point that with each toy will play at least one child, then there are definitely no more toys than children; In other words, instead of saying that each toy is a playing object for a child, we will say that the number of toys is not bigger than the number of children. These remarks allow in their turn the following natural statement: “we will say that the number of toys is the same with the number of kids if any two different children play with different toys and if with each toy plays a child.” All these expressions are in common language and present a set of conditions, in which we can say abut two groups of objects that they have the same number of objects (number of children and respectively number of toys in the case mentioned earlier). The description made is actually an explanation at the level of the common sense of some empiric evidence. In the above descriptions, all the words and propositions have clear significances which don’t cross, in the given situations, their limits of empiric significance.

Another common sense description in the common language is the following. Among the playing children there are boys and girls. We will naturally say that the group of children is enclosed in the group of playing children and that beside boys there are also girls in the group of children. If

all the playing children are for example from one playgroup and all children in this playgroup are actually those who play, we will naturally say that the two groups are actually the one and the same group of children.

Starting from these descriptions in the common language, let us build next in a formal mathematic language a few expressions and let us analyze the consequences. For instance: intuitively, a *set* is a collection of elements but the intuitive notion of a set leads to paradoxes, and there is considerable mathematical and philosophical disagreement about how best to refine the intuitive notion.[21] We will call the group of children "the set C" and that of toys the "set J?" The fact that a child c_i from the set C is playing only with one toy j_i from the set J, will call "function", which we represent as:[22]



$$f : C \to J, \quad f(c_i) = j_i$$

To say that two different children are playing with different toys and implicitly that there are no fewer toys than children, means mathematically [through synonymy[23]] that "the function f is injective". We can write this formally as $c_i \neq c_j \Rightarrow f(c_i) \neq f(c_j)$ or $c_i \neq c_j \Rightarrow j_i \neq j_j$ (or even $f(c_i) = f(c_j) \Rightarrow c_i = c_j$), which is no more than a mathematic

[21] It is of course assumed the preexistence of a mathematic vocabulary with the help of which the expressions can be built in the mathematic language.

[22] In order to be consequent with our initial affirmation that synonyms are words, we call "function" the expression: "relation child→toy". There are actually no difficulties in construction here. There are relations for which we have words. Synonymy as a relation will be between those words. We won't specify these details every time. It is a *formal* lexical condition in the last instance. For clarification, simplicity and for a clearer impression, the expression will be as natural as possible.

[23] This affirmation of synonymy becomes legitimate only from the perspective of rereading this part of the text after the whole analysis of the case of study.

expression, another reading of the proposition formulated above in the common language). To say that at one point with every toy a child plays and implicitly that there are no more toys than children becomes mathematically [through synonymy[16]] "the function f is surjective". Formally this can be written as $\forall j_i \in J, \exists c_i \in C$ so that $f(c_i) = j_i$ (emphasizing: which is nothing more than a mathematical expression, another reading of the proposition formulated above in the common language). To say that the number of children is equal with the number of players' results in saying again that there is a correspondence between child ↔ toy of one to one [through *synonymy*[16]] and that the function f is bijective " (both injective and surjective).

To say that in the set of the children C, beside boys (the set B) there are also girls (the set F) who play, results in saying, for example [through *synonymy*[16]] that the set B is strictly included in the set C. Set B is a *proper subset* of set C iff all the members of B are also members of C, but not all the members of C are members of B.[24] Formally this can be written as $B \subset C \Leftrightarrow \forall b_i \in B \Rightarrow b_i \in C, \exists c_j \in C, c_j \notin B$ (which is again a mathematic expression, a specific reading of the above formulated proposition in the common language). The *compliancy condition of the senses* means, consequently that the sense of the expressions in the two different languages: mathematic and common is given by the sense in the reference language, i.e. the common one. Starting from the common language, we will say [through *synonymy*] about two sets M and N that they are equal, formally M=N if they are formed from the same elements. Because this fact is not always visible from the beginning, the equality between sets is proven by the formal equivalence $M = N \Leftrightarrow M \subseteq N$ si $N \subseteq M$ which determines us again to say that all elements in the set M are among the elements of the set N and all elements in the set N are among the elements of the set M.[25]

---

[24] It follows from this definition that no set is a proper subset of itself.

[25] To define the applied symbols, we mention that the symbol "$\subseteq$" signifies the non-strict inclusion: $M \subseteq N \Leftrightarrow \forall m_i \in M \Rightarrow m_i \in N$ but the condition $\exists n_j \in N, n_j \notin M$ is not necessary.

We will summarize next the expressions [synonymies] built in the formal language, with the mention that "the words and propositions" are represented in the mathematical language and the senses are those from the common and reference language.[26]

Set: C, J, B, F, M, N.

Function: $f : C \to J, \quad f(c_i) = j_i .$

Injectivity (the injective function): $c_i \neq c_j \Rightarrow f(c_i) \neq f(c_j).$

Surjectivity (the surjective function): $\forall j_i \in J, \exists c_i \in C$ so that $f(c_i) = j_i$.

Bijectivity (the bijective function): injectivity and surjectivity

Strict inclusion: $B \subset C \Leftrightarrow \forall b_i \in B \Rightarrow b_i \in C, \exists c_j \in C, c_j \notin B .$

Equality of sets: $M = N \Leftrightarrow M \subseteq N$ and $N \subseteq M$.

Following the construction, we may affirm that the corresponding *synonyms* are lexical equivalences in order to express the same objects (the same referents) but in different languages.[27] It is what we called a form of "external synonymy". As a form of synonymy, it allows the words with identical meaning from different languages to be interchanged; they can alternate in a given linguistic "natural" or "mathematical" medium, without the global sense of the "natural" or "mathematical" message be changed. We can describe in this way the situation rendered in the drawing (D.3): "the children are playing bijectively with the toys". Of course it sounds unusual, but it is correct. We consider here as *function* the verb "to play", as *domain of definition* the group of children, as *range of the function* the group of toys and as *bijectivity* the *one-to-one correspondence*: children (member)/toy (member). So, two sets can be put into one-to-one correspondence iff their elements can be paired off such that each element

---

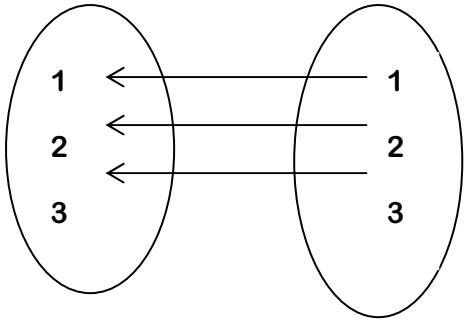
[26] The correspondence of the senses is obvious because we have only performed a mathematical formalization of some empiric situations, which are completely describable in the common language. The normal acceptance of the obvious things referring to the sense was one of the reasons for which the chosen examples were as simple as possible.

[27] Here is explicitly formulated the hypothesis present in the stages of the construction through the analogy of the formal correspondences: in each of the particular situations, the objectival reference was the same.

of the first set has exactly one counterpart element in the second set, and each element of the second set has exactly one counterpart element in the first set.[28] We don't normally express ourselves in this way, but that are cases of "fusions" where if the external synonymy is not broken, then the sense of the message remains the same.[29]

The presented example allows the following statement of *common sense*: "if between two sets a bijective correspondence (function) can be established, then the sets have the same number of elements". In mathematical words: "two sets have the *same cardinality* |A| = |B| iff they can be put into one-to-one correspondence or, if it can be constructed a bijective function $f : A \to B$; this definition applies, in mathematics, to finite as well as to infinite sets.".

Now, we will convene to use the word *intuition* in the following sense. On the one hand "our intuition makes us say that if all the elements of one set can be put in correspondence with all the elements of another set, then the two sets have the same number of elements". On the other hand, "if a one-to-one correspondence between the elements of a set and the elements of another set is possible, but for example the second set is strictly included in the first, then the second set has fewer elements then the first.



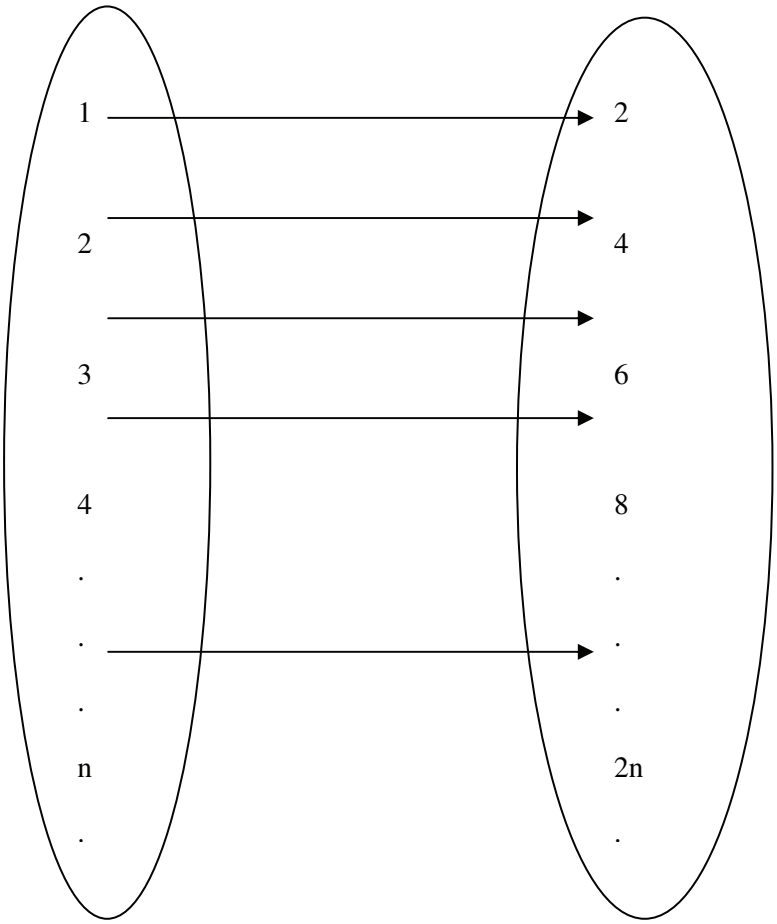
[28] We remind that in this case the synonyms are "words" but in different languages.

[29] The fusion of the common language with the mathematic one is not as frequent as in the case of the common language with the scientific ones (physical, biological and psychological). On the other hand, the mathematical language is present in the specialized scientific languages. These aspects are not considered in the present analysis. They are nevertheless important for the philosophical significance of the interferences and for the study of their consequences.

(This drawing simply illustrates the affirmed intuitive content.)

Let us describe the following in the mathematical language: the set of natural numbers N and the set of natural even numbers N_{2k}[30]. Putting two infinite sets into one-to-one correspondence is an infinite task and we cannot do it. We have an intuitive (empiric) support for this "action", step by step, but we cannot do it! So, to show that an infinite set like the even numbers can be put into one-to-one correspondence with an infinite set like natural numbers, we must "built" a mathematical strategy.

Let's consider the function $f : N \to N_{2k}$, $f(n) = 2n$, illustrated in the following drawing:



[30] For commodity we will consider the set of natural numbers $\mathbf{N}^*$, and implicitly the set of natural even numbers without zero.

N, N_{2k} - sets

$N_{2k} \subset N$ (for example $1 \in N$ dar $1 \notin N_{2k}$)

$f : N \to N_{2k}$, $f(n) = 2n$ - function

$n_i \neq n_j \Rightarrow 2n_i \neq 2n_j \Rightarrow f(n_i) \neq f(n_j)$ - injectivity (the injective function)

$\forall 2n_i \in N_{2k}\ \exists n_i \in N$ astfel ca $f(n_i) = 2n_i$ - surjectivity (the surjective function)

f injective and surjective function $\Leftrightarrow$ bijective function (bijectivity)

Remark: in the above demonstration regarding the existence of the bijectivity between N şi N_{2k} we had an intuitive support – without appealing to the axiomatic-mathematic frame – and mainly the fact that we understand which the successor of a number is: 4 is the natural successor of 3 and implicitly 4 is the even successor of 2, etc. In demonstrating the existence of a bijection between N and Q (the set of rational numbers $q = \frac{n_i}{n_j}, n_i, n_j \in N, n_j \neq 0$ [for simplicity we considered Q^*]), this intuitive support is not enough, because between any two natural consecutive numbers there is an infinity of rational numbers; furthermore, between any other two rational numbers there is an infinity of rational numbers; so that for a $q \in Q$ there is no proper successor. There are however demonstrations of the existence of a bijection between N and Q built on explicit intuitive supports. We mention here a terminological aspect: a set which can be put in a bijective correspondence with the set of natural numbers N is called "countable". (A set is "countable" iff its cardinality is either finite or equal to $\aleph_0$.) The sets: M={1, 2, 3}, N_{2k} and Q are countable. It is also said that the sets N, N_{2k}, Q have the same cardinal.[31]

There are demonstrations, for example that of the reductio ad absurdum also built on explicit intuitive supports, of the fact that the set of the real numbers R is not countable. This result has its importance: not all infinite sets are 'equally infinite'!

---

[31] The cardinal of the set of natural numbers is $|N| = \aleph_0$ (Aleph-null).

So, between the set N and the set N_{2k}, which is strictly included in the set N, we can build a bijection; in other words we can establish a correspondence of one-to-one! In other words, although the set of natural numbers has all the elements of the set of natural even numbers, and additionally a few other elements, the two sets can still be put in a correspondence of one-to-one, which make us return to the statement that they have the same number of elements! This last expression represents in the common language a comment on a mathematical result.

We will present in the following, still in the common language, a series of similar comments. How do we say? Is the set N "richer" in elements than the set N_{2k}? If by "richer" we understand *qualitatively also other elements* than the even numbers: the odd numbers, then yes indeed, the set is richer. Furthermore, if, as it has been mentioned, the set N contains all even elements, and beside these all odd elements, then doesn't "richer" imply not only *qualitatively- more elements* (elements of other nature, too) but is also *quantitative* "richer" - *more elements?* If not, then which is the impression created in the usual language by the following statement: "A set contains all elements of another set and in addition also other elements, but is as rich in elements as its subset"?! Or: "There are just as many even numbers as odd numbers but in the same time there are just as many odd numbers as even numbers in one place"?!

We will exemplify next a few other similar affirmations:

> […] a (bi)univocal correspondence between the set of positive integers and the subset of the even integers […] and then we naturally ask ourselves how can be these equalized?; a fact which argues against the familiar truth, the common sense, expressed through the thesis: :*the whole is bigger than any of its parts*[32]
>
> Il existe des "touts" aussi grands que certaines de leurs parties.[33]
>
> […] a unique one-to-one correspondence between the infinite list of the numbers and the infinite list of the even numbers. But all even

[32] Munteanu Marius, *Infinitul*; Presa Universitară Clujeană, Cluj-Napoca, 1999, p. 19.
[33] Verdier Norbert, *L´Infini en mathématiques*; Flammarion, 1997, p. 38.

numbers are included in the first list, against the fact that the "common sense" tells us that the even numbers must be only half of all numbers.[34]

We will focus in the following on the natural-intuitive relation *part/whole"*[35]. What can be said about a statement like: "it is a trivial truth that no finite set can be put into one-to-one correspondence with any of its proper subsets but the above example suggests that the part is equal with the whole"? (Every infinite set can be put into one-to-one correspondence with at least one of its proper subsets.[36]) This sentence, as it was presented, is neither paradoxal (logic) nor non-intuitive. This is a false proposition $N_{2k} = N \Leftrightarrow N_{2k} \subseteq N$ si $N \subseteq N_{2k}$ dar $N \not\subseteq N_{2k}$ [37]. The part is not equal to the whole. In this case, the correspondent of the word "part" from the common language can be through synonymy, in the mathematical language, the term (word) of "proper subset" built through the strict inclusion "$\subset$" and defined as follows: the set A is the proper subset of the set B if B contains all the elements of A but also other elements. But then, what can be said about the proposition: "The set of natural even numbers is not equal to the set of natural numbers, it is its proper subset (strict inclusion) and still, there are just as many natural numbers as natural even numbers!" How is it? Is it false? Is it senseless? We will see after a philosophical analysis of the language that this is mathematically no false proposition, and it is not senseless, it is simply non-intuitive at the level of the empiric intuition and implicitly at the level of the common language based on such intuitions.

---

[34] Barrow D. John, *Cartea infinitului*; Humanitas, Bucureşti, 2008, p. 65.

[35] The *part/whole* relation in its multiple metaphysic senses is not considered here, but it is also not excluded as a philosophic reference for a clearer image of the philosophic sense of the following considerations.

[36] We do not give here a proof for this theorem. A simple proof for it can be find, for exemple, in Suber Peter, *A Crash Course in the Mathematics of Infinite Sets*, St. John's Review, XLIV, 2 (1998) 35-59.

[37] $N \not\subseteq N_{2k}$ This non-inclusion can be easily demonstrated, but will be accepted here as a fact.

In the language used with the terms *qualitative-quantitative*, the strict inclusion $\subset$ implies a qualitative difference and the bijectivity a quantitative equivalence.[38] The remark which can be made is that, in certain mathematic situations, a qualitative difference introduced through strict inclusion may not affect the quantitative equivalence kept through bijectivity. This is the reason for which, in defining the "proper subset": "The set A is the proper subset of the set B if B contains all elements of A but contains also other elements" ,we have avoided the "The set A is the proper subset of the set B if B contains all elements of A but contains *additionally* also other elements": How do we say then: "The part is not equal to the whole but the whole is not more than the part!?[39] Answering in linguistic elements, to the question "How do we say?" wanted to suggest the existence of a "tension" in the common language caused by the appearance of some "intuitive paradoxes".[40] On the other hand, it comes naturally to ask: "Are these examples difficulties or *problems* at the level of the propositional sense?" An attempt to clarify exclusively from the perspective of the language doesn't seem to be possible. Other philosophic directions will be approached in parallel.

For the beginning, let us consider the following preparatory description. As it has already been suggested, "the quantity"- "the number" of elements of a set is called mathematically the *cardinal* of a set. The *cardinality* of a set is the number of members it contains. The cardinality of set M is |M|. Hence while M is a set, |M| is a number. A finite set M with n elements has the cardinal $|M| = n$ (a [finite] natural number). An infinite set,

---

[38] This distinction of the language deserves a special attention; nevertheless, this won't be developed in the current analysis.

[39] The distinction *part-whole* also deserves a special attention; nevertheless, this won't be developed in this context.

[40] The collocation "intuitive paradoxes" was introduced here through analogy with the logic paradoxes or logic-mathematic paradoxes, just to emphasize the existence of an "intuitive tension" at the language level. We illustrate through another example of the so-called *intuitive tension* at the language level, the reference to the negative numbers fro the next fragment. "In Europe there were some difficulties in accepting the notion "less then nothing" -Guedj Denis, *Matematica explicată fiicelor mele* Transl.Alexandru Siclovan, Publishing House Cartier, Bucureşti 2008, p. 45

as the set of natural numbers *N* has, by definition, the cardinal $|N| = \aleph_0$ (a *transfinite number*; a *transfinite number* or *transfinite cardinal* is the cardinality of some infinite set).[41] There are infinite sets larger than the set of natural numbers: an infinite set as the set of real numbers *R* has the cardinal $|R| = \aleph_1$, another transfinite number and $\aleph_1 > \aleph_0$ (a set is *uncountable* iff its cardinality is greater than □0, then *R* is uncountable).[42] Let the following mathematic propositions be:

1) $A \subset B \Rightarrow |A| < |B|$.

2) $A \subset B \Rightarrow |A| = |B| \text{ sau } |A| < |B|$.

Two questions arise: a) "can the proposition 1) become a proposition with a factual significance?"[43] But, b) "can the proposition 2) become a proposition with a factual significance?"

Let there be two suggestions for answers. Answer a) Yes, the proposition 1) can be brought in the common language with an empiric significance: "The set B contains all elements of the set A and additional other elements and is richer in elements than the set A [for finite sets]" (from the general mathematic point of view, this proposition is false

---

[41] |S| is a cardinal number as opposed to an ordinal number. A cardinal number answers the question "*how many?*" but an ordinal number answers the question "*which one?*". Linguistic, natural numbers are used both ways in different contexts (one/the first; two/the second). Sets have cardinality, they have some definite number of members but they do not have ordinality, their members are not in any particular order.

[42] In this analysis of synonymy we ignore the mathematics of infinite ordinals and we consider only the mathematics of infinite cardinal numbers.

[43] Through factual significance we understand here a general aspect, usually the possibility for some symbols to have an empiric reference. The factual significance implies also the possibility of a factual sense, i.e the possibility to "express" the proposition in a common or scientific language, where the way of exposing the reference has a sense specific to the corresponding language. For example in the mathematic relation $\vec{F}(\vec{r}(t)) = m\frac{d^2\vec{r}(t)}{dt^2}$, *m* is the mass of a mechanic object, $\vec{r}$ is its position, *t* is the *time* measured with a clock associated to the dynamics of the object and the mathematic expression has a physical meaning, representing Newton's Second Law in motion. We point out that in the text is not considered a too restrictive factual reference, as in the above example, in which the symbols are associated with physical measures. Subsequently, the *finitude* can be a factual reference.

because, if the sets are infinite, the equality of cardinals in the given conditions is possible: $A \subset B$, $[B \setminus A \neq \emptyset]$) Answer b)."Transferring the proposition 2) in the common language with empiric significance is only partially possible. Partially means that only the second part of the implication, after *or*, can have logically empiric significance. This thing is not mandatory because *A* and *B* can be infinite $N \subset R$, $|N| = \aleph_0 < \aleph_1 = |R|$, but it is possible." A special remark, which we might make, is that 1) totally satisfies our intuition, while 2) satisfies it partially. An important addition: the set of the natural numbers is a mathematic object different from its finite subsets, the only ones in correspondence with empiric objects.

We remind here that we brought the empiric to finitude[44] and that for the one-to-one correspondence between the elements of two sets; we built and used the mathematical term bijectivity. What has been actually done in the entire "construction"? We have passed through the relation of synonymy between words belonging to different languages, from descriptions in the common language to descriptions in the mathematic language? *Synonyms* have as "semantic reference" in their construction, in the last instance – the same basic *object*: *the finite sets* (including in their constructions, with reference to this *object* also the presented relations (the object remained well defined every time). For example, this is how we passed from the correspondence of one-to-one child↔toy to the bijectivity of the function $f : C \to J$, $f(c_i) = j_i$. As long as the "mathematic objects" have the "finitude" in common with the "empiric objects", there are no 'problems', actually the relation of synonymy is possible. The mathematic language built in this way through synonymies. The mathematic language built in this way through synonymies extends its coverage also over other "objects" specific to the mathematics, like the infinite set of the natural numbers or of the natural even numbers, which, through the quality of their infinity, don't have an empiric correspondent.[45]?! On the one hand, through this extension of the "coverage" over some "objects without an empiric

[44] More exactly, the *finitude* can be an empiric objectival reference.
[45] For $\forall n \in N, \exists n+1 \in N, n+1 > n$.

correspondent", the mathematic words (strict inclusion, injectivity, surjectivity, etc.) built as synonyms become homonyms in the sense in which the "semantic reference" (the referent) change! On the other hand, through this extension of "coverage" over some "objects without an empiric correspondent" the given terms (strict inclusion, injectivity, surjectivity, etc), more generally the mathematic language doesn't overcome its limits, doesn't enter in logic or terminological contradiction. In other words, the mathematic language built here starting from the common language which serves to the empiric, is used net to describe other types of sheer mathematic "objects' (In a mathematic expression, the referent changes). The results of the "mathematic speech" over these kinds of objects are mathematic propositions, which don't always change into propositions with empiric significance, i.e. propositions with a factual sense in the common language. Furthermore, their transfer in the common language, apparently through the usage of synonyms, leads to expressions of the type "the part is equal to the whole" – "intuitive paradox propositions". Other expressions of the type "The set of all natural numbers and the set of just the natural even numbers are equally rich in elements" are beside non-intuitive and evasive also ambiguous. Moreover, we might say that propositions like: "There are just as many natural numbers as the natural even numbers." – accepted as a proposition of the common language doesn't have a significance, resp. it has no sense (according to the initial hypotheses, the problem of sense is not an issue in this case). Convening that this is a mathematic proposition with "natural terms" which expresses the fact that between N and N_{2k} can be established a bijection – i.e it is mathematic possible that $A \subset B$ and $|A| = |B|$ - then this has a mathematic significance, it is simply a mathematic result.[46] Thus, we can obtain mathematic results, which the common language with reference to "its objects" cannot express. To say that $A \subset B$ and still a bijection can ne found $f : A \to B$, that is to say that A

[46] The proposition can be reformulated with the help of the mathematic notion of "cardinal" in this way: "The set of the natural numbers and the set of natural even numbers have the same cardinal: $|N| = |N_{2k}|$" But these types of details are not necessary for the purpose of this analysis, although they would offer us some additional clarifications.

and *B* have the same cardinal $|A| = |B|$ is a "conceptual result" and from the mathematic point of view is nothing "wrong', "paradoxal intuitive", "unusual', etc. Only by trying to express this mathematic result in the common language, "irregularities" and "intuitive paradoxes" appear.

Questions: What will we say? That the sense of the proposition didn't actually change, it is just non-intuitive? Or that the sense of the proposition didn't change? Or even more, that the proposition has no sense? Or furthermore, that we don't even have a proposition in the form asked by the propositions of the common language? Is in this case the following statement legitimate: "if in the construction of external synonyms the basic proposition is being violated[47], here, the construction of the semantic reference and the constructed synonyms don't refer anymore to the same object [referent], then we can come to a breakage *of the linguistic symmetry*, i.e. in the new propositions the propositional sense is not conserved anymore. Could this breakage of symmetry embrace the form of the *intuitive paradox?"* If we remain, on the one hand within the frames imposed by what we called the defining features of the synonyms, and on the other hand we remain consequent to an "axiomatic vision" of the analysis, the answer is obvious: the terms were synonyms only as long as they referred to the same object, had the same semantic reference (the same referent). When the semantic reference changes (or disappears (!)), any reference to the synonymy, for example the problem of preserving the propositional sense, is not founded. We talk about different things, we have different propositions. Is the appearance of the intuitive aspects a problem of propositional sense? They are of course subjects of real philosophic interest but are intuitions also an issue of language? The present analysis promotes generally the study of a possible philosophic relation *intuition/language* through the analysis of the particular philosophic case *intuition* and *synonymy***.**

Before trying to answer the previous questions, let us continue briefly with a few considerations over "intuition" and "infinite'. As a mathematic notion, "the infinite" is defined precisely in varied mathematic

---

[47] "The synonymy is a type of relation between words referring to the same object."

domains and an "object" such as "the infinite set" is completely integrated in the mathematic language. Actually it has an ontical statute only *in* and *through* the mathematic language.[48] Without defining here *the intuition* we will convene to say that the notion of "infinity" has in the common language an intuitive content and metaphoric representations: "endless", "immeasurable", "limitless" , "boundless", "illimitable", etc. a definition for 'intuition" is problematic. The usage of the word *intuition* in different situations and examples allows somehow 'indirect' characterizations of it. Let be the following fragment whose language is as common as possible and whose message is a pedagogic one, of clarification at the elementary mathematic level.

"Everyone supposes that the series of natural numbers doesn't stop, we can always go further. There is no "biggest natural number'. If there were any, let us call him *A, A* being natural and $A+1$ would be a similar one. Either $A+1$ is bigger than *A*, which contradicts the fact that *A* is the biggest. This kind of demonstration is a *demonstration of reductio ad absurdum* so, there is no natural number - the biggest one, which implies the fact that it is an infinite quantity."[49]

Let us analyze the fragment. It is affirmed as intuitive the impression or the conviction, that the series of the natural numbers can always be completed, continued. A basis for this expression (or conviction) is the every day experience. We have a quantity of objects and we can expand it by adding an extra object. The common sense tells us that we can o this. If not effectively every time, we can at least imagine the steps. At the "mental" level, two presuppositions, not necessarily explicit, are present: the first, that nothing stops the mind to make additions; the second that through the addition something bigger or more is obtained. Remaining strictly within the frames of these considerations, saying that there is a biggest finite quantity, where we can not make any addition, is something non-intuitive, which

---

[48] We mention that in the current analysis only the natural-empiric and the formal-mathematic elements are considered, any other metaphysic or speculative logic-philosophic approach is, as far as possible avoided.
[49] Guedj, 2008, p. 23.

"contradicts" the intuition, contradicts the good sense, is something unusual. We accept the fact that if something is non-intuitive is not necessary irrational at the same time; or that a contradicted intuition is not a demonstration of reductio ad absurdum. We won't discuss here the axiomatic of the natural numbers. We will approach here only aspects of *intuitive* interest. Mathematically, if *A* is a natural number, the fact that $A+1$ is on the one hand a natural number, and on the other hand it is bigger (or more) than *A*, it is a quantitative successor, with other words it should be postulated. There are sets of numbers in which there is no relation of order and there are classifications which do not refer to quantitative contents.[50] From the strict mathematic point of view, the postulates can be or not neutral (the case of the modern mathematics) towards an intuitive support (content). The well-known example of the Euclidian geometry and of the non-Euclidian geometries is perhaps the simplest illustration of this aspect. In the quoted text, *the intuitive level and the rational level* are overlapped. This interference, however, at the level of the clarifying suggestion, is a first stage in explaining and is an imminent, common procedure, Many times, intuitive contents are intrinsically presented in the rational discursiveness. What we wanted to emphasize through the reference to this chapter is the fact that we can try explicit rational explanations on intuitive implicit backgrounds. The present analysis is also an example in this sense.

Coming back,

> The equivalence of the infinite sets shows us at what examples we can expect in the arithmetic field of the infinite and that the common sense cannot comply with the infinite.[51]

This however, doesn't stop the legitimacy of a question of the type: "Is there a possibility for a mathematic notion such as *the infinite (*"∞"*)* to be found out in one way or another, if not a proper "physical object", then at

---

[50] For example, the set **C** of the complex number, respectively the lexicographic order.
[51] Munteanu, 1999, p. 19

least a factual situation which should (empirically) signify it?"[52] We will accept some evidences imposed by the current usage of the language. So that, the *word* "finite set" has a clear significance in the common language, being in univocal correspondence with a well defined given object, even if the objectival-factual referents are multiple: different groups of certain objects from the surrounding world. On the other hand, the notion of "finite set" also has a clear significance in the mathematic language, being in a univocal correspondence with a well defined given object, even if it doesn't have objectival-factual referents. We assume that at this level of analysis, the "natural" thinking and the "mathematic" thinking are unitary in the sense that the object chosen for the "finite set" is the same. The common language, with reference to the "empiric object' – well definable – *the finite set*, says something with sense, intelligible and with coverage in the empiric intuition. Built through *synonymy*, starting from the same "empiric object" *finite set*, the mathematic propositions say in other words the same thing with sense, intelligible and with an intuitive coverage. Actually, there are used different symbols, specific to each language, in order to designate the same object [referent].[53] If the mathematic language passes beyond these empiric referential limits *finite sets* and with the same specific words *surjectivity, injectivity, bijectivity, cardinality* it talks from its own point of view, of "new mathematic objects" *infinite sets*, then its intelligibility, guaranteed by the intrinsic rationality is being preserved. The factual reference – objects, properties, identifiable in the field of the empiric – doesn't exist anymore. The sense of the mathematic propositions is now

[52] There are mathematic notions which at the level of the common sense seem "less intuitive" than the one of "infinite" as it is for example the notion of "fractal dimension" for which we can difficult find metaphoric representations – "how could we imagine the dimension of an *object* situated between dot and line, between line and plan? – but for which there are factual contexts of significance (physical situations characterized by chaos and complexity)

[53] There is another tacit acceptance which is a fact: on one hand, as abstract as a language, formally built, may be or as specialized a scientific language may be, their sources are in the common language.

pure mathematic, 'unnatural".[54] For instance, in this mathematical context, *the infinite sets*, and only them (all of them), possess the non-intuitive property that they can be put into one-to-one correspondence with at least one of their proper subsets and this property is a *necessary and sufficient* condition for being an infinite set. We can start and define an infinite set in this way but we have not an intuitive support for the definition. Our empiric intuition is being "puzzled".

Let us briefly present in the end of the case of study, through a suggestion of response to the questions formulated above regarding the propositional sense from the common language, the following observations. There are mainly *two possibilities:*

1) "The infinite" is not an objectival reference included in the common language. In this way there is no o in $((c, o), p)$ naturally definable and so, according to the premises, from the perspective of the common language where the expression is being made, the respective propositions are not legitimate expressions of the language, they are actually no propositions. There is no authentic "tension" in this language. There are no difficulties or problems at the level of the propositional sense. The so-called "intuitive paradoxes' are forms of interpretation of some non-propositional linguistic constructions and as a result they have no object or sense. The semantic reference has been lost; consequently, any reference to synonymy is invalid. In this way we can accept that the appearance of some non-intuitive objects is a signal of existence of some problems of language. This first possibility of answer is quite a radical one, more theoretical and it encourages the analysis of a possible critical philosophic-linguistic relation *intuition/language.*

2) "The infinite" is an objectival reference expressed in the common language, with the remark about this reference that it is *intuitive* (in the variant suggested by the metaphoric representations rendered previously) and it is still not well-defined. Let us accept on the one hand, that the object

---

[54] Once again, this expression doesn't mean at all that mathematics, through its own concepts, without a possible empiric significance in the sense described above, cannot be described in the physical world.

of reference - *the infinite* - is the same "mathematically" and "naturally" and that consequently, from the mathematic and natural point of view, we talk about "the same thing". In this way, the first condition of synonymy, the one of the objectival semantic reference of the identity, is not being infringed. We accept on the other hand that in the given conditions: • the mathematic notion of infinite has its origins in the current natural intuitions; • the mathematic concepts used in the mathematic description are the same with those used for finite sets; • "the construction" of the mathematic concepts used in the description of the finite sets was made in direct correspondence with similar descriptions from the common language; so, in these conditions, we may assume that by talking about the infinite in the common language "we talk approximately about the same thing as in mathematics, but with other words." Furthermore, we may sustain that the sense of the propositions (as a means of exposing the reference) from the common language through the introduction of the mathematic propositions in this language, is the same with the mathematic sense. There is, in other words a correspondence of the senses. It is also assumed that the statements remain valid also for the propositions directly derived in a way or another from this language. In these conditions, the following question naturally derives: "How are the *intuitive paradoxes* created?" The response depends on more factors. Two factors seem to be basic: a) we express mathematical subjects in the common language; b) there is a 'linguistic pragmatism", a usual, current, daily way of talking in which the reference objects are empirical – the finite sets is such an example – and the senses of the propositions are exposures of some of these referents. The intuitive paradoxes appear as a result of the interference of these basic factors. A natural and simple expression of the interference is the following: "we talk about infinite sets but we think about finite sets." We legitimately talk about infinite sets, because we convened that this is possible, but in fact, the reference object of those mentioned are still the finite sets. If we accept that we express in usual words mathematic results on mathematic objects, then we accept that we only make a "mathematic exercise" in a mathematic linguistic form and this

is it.[55] There is not even a combination of languages in the sense of the mentioned proposition: "the children are playing bijetively with the toys", it is about a totally natural expression of some mathematic situations. What is implicitly being done, is the fact that what the common language says about the infinite sets (an object) is being compared with what the common language says about the finite sets (*another* object). It might be affirmed that the propositions have different senses, that a breakage of the linguistic symmetry is caused, i.e. that in the new propositions the propositional sense is not being preserved anymore. But through the presuppositions made at the beginning, it is obvious that the propositions have different senses, because they refer to different objects. From the perspective of the relation of synonymy, as it has been characterized, it can be said that *the same words* are not synonyms. Let us resume in the end these aspects. There, where, tacit, implicitly a relation of synonymy, which actually doesn't exist because of the difference of objectival reference (in other words, because of the semantic difference of referentiality) is being assumed, "intuitive paradoxes" can appear. Consequently, this *second possibility* leads in final instance to the conclusions of the *first possibility*: the semantic reference has changed and as a result any reference to the synonymy is invalid. Thus, it can be accepted that the appearance of some (non)intuitive aspects is a sign of existence of some problems of inadequate usage of the language. This second possibility of answer can be viewed as a completion but also as an emphasis of the first. In its turn, this second possibility encourages the analysis of a possible philosophic- linguistic critical relation *intuition / language* but from the perspective of a form of linguistic pragmatism, of inadequate usage in the possibility mentioned here, of the language.

[55] There is also a reciprocal "phenomenon". It is known that at least in the stages of initiations in mathematics (and not only), in the first steps of familiarization with the mathematic formal language; we have tried here to charge the mathematic abstract symbols with intuitive natural contents taken over from the common language. This supports greatly the "mathematic understanding". No expansion of this subject will be made.

References

Barnes W. Donald & Mack J. M., *An Algebraic Introduction to Mathematical Logic*, Hardcover, Springer, 1975

Barrow D. John, *Cartea infinitului*; Humanitas, Bucureşti, 2008

Bulgăr Gh, *Dicţionar de sinonime*; Editura Lucman, 2004, Bucureşti

Deely John, *Bazele semioticii*; Transl. Mariana Neț, Editura All, Bucureşti, 1997

Guedj Denis, *Matematica explicată fiicelor mele* Transl.Alexandru Siclovan, Publishing House Cartier, Bucureşti 2008

M. Lynne Murphy**,** *Semantic Relations and the Lexicon: Antonymy, Synonymy, and Other Paradigms*, Cambridge University Press, 2003

Mac Lane Saunders, *Categories for the Working Mathematician*, Springer, 1998

Moore A. W., *The infinite*, Routledge, 2001

Munteanu Marius, *Infinitul*, Presa Universitară Clujeană, Cluj-Napoca, 1999

Serre Jean-Pierre, A *Course in Arithmetic*, Springer-Verlang, 1978

Suber Peter, A *Crash Course in the Mathematics of Infinite Sets*, St. John's Review, XLIV, 2 (1998), 35-59.

Takeuti Gaisi & Zaring M. Wilson, *Introduction to Axiomatic Set Theory*, Hardcover, Springer-Verlang, 1982

Verdier Norbert, *L´Infini en mathématiques*; Flammarion

Wittgenstein Ludwig, *Tractatus logico-philosophicus*; Transl. Dumitru M. & Flonta M., Editura Humanitas, Bucureşti, 1997

On the effectiveness of Kalmár's completeness proof for propositional calculus

Adrian LUDUŞAN [*]
Babes-Bolyai University Cluj-Napoca

Abstract:

Ever since Kurt Gödel's proof of the completeness theorem of first-order logic in 1930 other few alternative proofs have been produced, whose logical, mathematical or epistemological virtues are worth taking into consideration. In what follows we will deal with one of these alternative proofs for propositional calculus, namely that of Laszlo Kalmár. What strikes as remarkable in the case of this proof is, on the one hand, its constructive character, which offers an effective procedure of determining the proof of any tautology within the respective propositional calculus, and on the other hand, its simplicity. In his completeness proof, Kalmár uses a crucial lemma which glues syntactical derivation with semantic computation. The aim of this paper is to highlight two ways of understanding the effectiveness of Kalmár's proof for this lemma, and to pinpoint a small problem regarding the effective character of the lemma alongside a solution to this problem.

Keywords: completeness theorem, propositional calculus, completeness proof, Kalmár's lemma, effective procedure, effective proof.

I. Preliminaries

Ever since Gödel's proof of the completeness theorem[1] of first-order logic in 1930 other few alternative proofs have been produced, whose logical, mathematical or epistemological virtues are worth taking into consideration. Without doubt, the most famous alternative proof was

[*] E-mail: adrianludusan@yahoo.com.
[1] Kurt Gödel [1].

elaborated by Leon Henkin in his article published in 1949[2]. The virtues of this proof go beyond simple epistemological considerations regarding its elegancy and simplicity, bringing forward a new and ingenious way of constructing the model of a theory, respectively to use the syntax of the theory as raw material for constructing the model. In 1959 Jaako Hintikka[3] proposed a proof related to Henkin's, in which he assumed the construction of a set of formulas having certain properties which allow a "natural" identification of a model of the respective formulas and from this result, a proof of the completeness theorem. Meanwhile, Hintikka's proof became the "standard" proof theorem for first order logic via semantic tableaux, and natural deduction systems [4]. Together the two types of proof became "canonical", most textbooks on logic giving one of them as proof of the completeness theorem.

In this period we can also witness a mathematical "recovery" of the completeness theorems on two major lines: algebraic and topological proofs[5] of completeness for first-order logic. What this recovery did, on the one hand, was to set up a correspondence between the logical results of completeness and certain mathematical results, such as Stone's representation theorems[6] or Tychonoff's theorem, and, on the other hand, to reveal and give an epistemic awareness to previous results, the most notable example being the compactness theorem of first-order logic[7].

Historically speaking though, Gödel's proof wasn't the first proof of completeness of a logical system. Already in 1921, Emil Post[8] had given the first proof of completeness, but for a narrower system of first-order logic, namely for propositional calculus, as it was formulated in "Principia

---

[2] Leon Henkin [2].
[3] Jaakko Hintikka [3].
[4] see Ian Chiswell and Wilfrid Hodges [4].
[5] For the topological proofs of completeness see: Evert W. Beth [5], and for the algebraic proofs see Jerzy Łos [6] or Helena Rasiowa, Roman Sikorski [7].
[6] Marshall H. Stone [8].
[7] For the difficult process of understanding the epistemological relevancy of the compactness theorem and its connection with the Löwenheim–Skolem theorem and the completeness theorem see John W. Dawson [9].
[8] Emil Post [10].

Mathematica" by Russell and Whitehead[9]. What is curious is that the alternative proofs for the completeness of propositional calculus have emerged after Gödel's proof and not after Post's. In what follows we will deal with one of these alternative proofs, which also appeared after Gödel's, namely that of László Kalmár.[10] What strikes as remarkable in the case of this proof is, on the one hand, its constructive character, which offers an effective procedure of determining the proof of any tautology within the respective propositional calculus, and on the other hand, its simplicity. For instance, Alonzo Church, exposing the proof method of the completeness theorem for propositional calculus that he employed in his book, *Introduction to Mathematical Logic*, says that:

> the idea of applying Kalmár's method to this formulation of propositional calculus [the one present in § 10 of *Introduction to Mathematical Logic* n.n] was suggested to the writer by Leon Henkin as yelding perhaps the briefest available completeness proof for the propositional calculus (if based on independent axioms with *modus ponens* and substitution as rules of inference)[11]

Moreover, in an article in which he generalizes Kalmár's method of proof in order to obtain a complete axiomatization of any fragment of propositional logic which includes implication among its connectives, Henkin himself states that:

> Of the several methods for proving the completeness of sets of axioms for the propositional calculus perhaps the simplest is due to Kalmár, although it does not appear to be widely known.[12]

More precisely, our paper will focus on the proof of a lemma, necessary in establishing a Kalmár-type proof of completeness. This lemma

---

[9] Bertrand Russell, Alfred Whitehead [11].
[10] László Kalmár [12].
[11] Alonzo Church [13], p. 163, n. 288.
[12] Leon Henkin [14], p 42.

plays the most important role in the proof of the completeness theorem, because it sets up a correspondence between the system's syntax and semantics. This function of Kalmár's lemma, with the importance deriving from it, can also be found in the aforementioned completeness proofs, namely those of Hintikka and of Henkin. Both proofs are based on a lemma which sets up a correspondence between the system's syntax and semantics. Due to this reason this type of lemmas received a suggestive technical denomination: model existence lemma.

For a clear understanding of what the lemma asserts, we must make a few specifications regarding the propositional calculus system ($\mathbf{C_p}$) for which we will determine the lemma. To be more precise, we will discuss the syntactic and semantic aspects relevant for the understanding of the lemma.

II. Setting the stage

II. A. A formal system of propositional calculus ($\mathbf{C_p}$)

The formal system of propositional calculus is the quadruple of sets $\mathbf{C_p} = \{\mathbf{A}, \mathbf{F_p}, \mathbf{Ax}, \mathbf{R_d}\}$ in which:

1. A is the set of the system's alphabet

2. $\mathbf{F_p}$ is the set of the system's formulas

3. Ax is the set of the system's axioms

4. $\mathbf{R_d}$ is the set of the system's rules of derivation.

1. In our case, the system's alphabet is formed out of the set of propositional variables, the set of logical connectives and the set of the punctuation symbols, more precisely:

a) the countable set of propositional variables, $\mathrm{Var} = \{p_n: n \in N\}$

b) the sentential connectives, $O = \{\rightarrow, \neg\}$

c) the punctuation symbols, $P = \{(,)\}$.

The **alphabet** of propositional calculus $\mathbf{C_p}$ is formed, in this case, out of the reunion of all the sets from a) to c), that is:

$\mathbf{A} = \{\mathrm{Var} \cup O \cup P\}$.

Intuitively, propositional variables stand for simple[13] sentences liable of truth value, the two logical connectives $\rightarrow$, $\neg$ stand, *cum grano salis*, for "if - then" and "not" as used in natural language and the punctuation symbols are the parentheses (left and right), used to render a unique reading of the formulas. With the alphabet A we can form the set A^* of words over that alphabet, defined as the set of all finite strings which can be formed using symbols from A. Algebraically, the set A^* with the concatenation operation acting on it forms a freely generated subgroup. From the set of all finite strings that can be formed using the symbols of the alphabet A, we distinguish a class of strings, called well formed formulas, inductively defined as it follows:

2. Definition (well formed formula –wff): the set $\mathbf{F_p}$ of well formed formulas (wffs) of $\mathbf{C_p}$ is the smallest set which satisfies the following clauses:

1. If $p_i \in$ Var then $p_i \in \mathbf{F_p}$
2. If $x \in \mathbf{F_p}$ then $(\neg x) \in \mathbf{F_p}$
3. If x and $y \in \mathbf{F_p}$ then $(x \rightarrow y) \in \mathbf{F_p}$
4. A finite string of symbols from A is a wff if and only if it can be obtained by applying a finite number of times the foregoing clauses 1. – 3.

The way in which the set $\mathbf{F_p}$ was constructed allows us to use two powerful mathematical instruments, essential in the development of propositional calculus and the establishing of relevant results for the characterization of propositional calculus: the method of proof using induction and the method of defining functions using recursivity. Using the aforementioned techniques of induction and recursivity we can prove the following theorem, essential for our further analysis:

Unique Decomposition Theorem: any formula $x \in \mathbf{F_p}$ has exactly one of the following forms:

a. $x \in$ Var

b. $x = (\neg y)$, where y is a uniquely determined formula

c. $x = (y \rightarrow z)$, where y, z are uniquely determined formulas.

---

[13] Simple is to be understood as indecomposable.

Having established the Unique Decomposition Theorem we can further characterize the notion of complexity of a formula and of immediate subformula and subformulas of a given wff x.

As we have mentioned above, a useful notion we will use through out the subsequent sections is the notion of complexity c(x) of a wff x, which we will define based on the number of occurences of logical connectives, following the same method we have used when we defined the set $\mathbf{F_p}$:

3. Definition (complexity of a wff)

1. If $x \in$ Var then $c(x) = 0$
2. If $c(x) = n$ then $c(\neg x) = n + 1$
3. If $c(x) = n$ and if $c(y) = m$ then $c(x \rightarrow y) = n + m + 1$.

Definition (immediate subformula): let $x \in \mathbf{F_p}$

1. if $x = p_i$ then x has no immediate subformulas
2. if $x = (\neg y)$ then the only immediate subformula of x is y
3. if $x = (y \rightarrow z)$ then the only immedite subformulas of x are y and z.

Definition (subformulas – subform(x)): let $x \in \mathbf{F_p}$.

1. if $x \in$ Var then subform(x) = {x}
2. if $x = (\neg y)$ then subform(x) = subform(y) $\cup$ {x}
3. if $x = (y \rightarrow z)$ then subform(x) = subform(y) $\cup$ subform(z) $\cup$ {x}.

3. Axioms (Ax):

Ax_1: $(x \rightarrow (x \rightarrow y))$

Ax_2: $((x \rightarrow (y \rightarrow z)) \rightarrow ((x \rightarrow y) \rightarrow (x \rightarrow z)))$

Ax_3: $((\neg x) \rightarrow (\neg y)) \rightarrow (y \rightarrow x))$

4. Derivation rules (R_d):

Modus ponens (mp): if x and y are wffs, than from x and $(x \rightarrow y)$ we infer y.

Formally, $x, (x \rightarrow y) \vdash y$

Observation: as we can notice both in the above definitions and in the specification of axioms and inferring rules, we have used certain variables which do not belong to the alphabet of the specified system. Technically speaking, these are metavariables. We assume the distinction between language and metalanguage and therefore the distinction between variables and metavariables is familiar to the reader, as is the distinction between axioms and axiom schemata. According to the aforementioned, the axioms $\mathbf{Ax_1 - Ax_3}$ are axiom schemata. Also we will not insist on the way the substitution rule is replaced by the way the axiom schemata functions.

One of the reasons for which we construct a formal system is the precise definition of what exactly constitutes a proof.

Definition (Proof)

In the considered axiomatic system, **a proof** is a finite string of wffs x_1 - x_n such that $x = x_n$ and for any $k \leq n$,

1. $x_k \in$ **Ax** (x_k is an axiom)

or

2. x_i, x_j $\vdash x_k$ (x_k is inferred from x_i, x_j, according to the derivation rule modus ponens, where $i, j < k$,)

The string x_1 - x_n is the proof of the wff x. A wff x is called a **theorem** of an axiomatic system if there is a proof of it in that system. Formally: $\vdash x$

In what follows, by a formula we will understand a well formed formula. It is convenient to introduce the notion of proof by assumptions or by hypotheses, for a clear understanding of our discussion. By deduction we will understand such a proof made through assumptions. Let Σ be a set of formulas of the considered calculus, where the elements of Σ are assumptions or hypotheses.

Definition(deducibility)

We call **deducibility** of a formula x from the assumptions or hypotheses Σ a finite string of formulas x_1 - x_n such that $x = x_n$ and for every $k \leq n$,

1. $x_k \in$ **Ax** (x_k is an axiom)
2. $x_k \in \Sigma$ (x_k is one of the formulas of Σ)
3. $x_i, x_j \vdash x_k$ (x_k is inferred from x_i, x_j, according to the derivation rule modus ponens, where $i, j < k$,)

The string x_1 - x_n is the **deduction** of formula x from the assumptions or hypotheses of Σ. If there is such a deduction of a formula x from Σ we say that x is **deducible** from Σ Formally: $\Sigma \vdash x$.

In what follows we shall enumerate without proving some results necessary in setting up a Kalmár-type completeness proof of propositional calculus, namely a few theorems of $\mathbf{C_p}$, some properties of the deducibility $\vdash$ relation as well as a theorem which establishes a correspondence between the implication $\rightarrow$ and the deducibility $\vdash$ relationship.

Theorems of C_p:

1. $\vdash (x \rightarrow (\neg(\neg x)))$
2. $\vdash ((\neg x) \rightarrow (x \rightarrow y))$
3. $\vdash (y \rightarrow (x \rightarrow y))$
4. $\vdash (x \rightarrow ((\neg y) \rightarrow (\neg(x \rightarrow y))))$
5. $\vdash ((x \rightarrow y) \rightarrow (((\neg x) \rightarrow y) \rightarrow y)$

Properties of $\vdash$:

1. $x \vdash x$ (the assumption property –AS.)
2. if $\Sigma \vdash x$ and $x \vdash y$ then $\Sigma \vdash y$ (the cutting property – CUT)
3. if $\Sigma \vdash x$ then $\Sigma \cup \Delta \vdash x$ (the thining property - THIN)

Deduction theorem: if $\Sigma, x \vdash y$ then $\Sigma \vdash (x \rightarrow y)$

II. B Semantics:

In order to construct the semantics of a formal system we must identify a systematic way of attaching meaning to the syntactical elements starting with the atomic components, in our case, the propositional variables. For the propositional calculus we have developed so far, we shall do this using valuation functions and their extensions.

Definition (valuation):

A **valuation** is a function v which assigns a definite truth value to each propositional variables (in this case either false – 0, or true – 1), v: Var $\rightarrow \{0,1\}$. The set of all possible valuation functions v is: $2^{\overline{Var}} = 2^{\aleph_0} = c$, where $\overline{Var}$ is the cardinal number of the set of propositional variables Var. Obviously, in the case of a formula $x \in \mathbf{F_p}$ composed of the propositional variables $\{p_1, \ldots, p_n\}$, case which we will abbreviate as $x(p_1, \ldots, p_n)$, we have 2^n possible valuations.

Definition (interpretation):

An interpretation i is the unique extension[14] of a valuation function v, over the set $\mathbf{F_p}$ of formulas i.e $i = \bar{v}: \mathbf{F_p} \rightarrow \{0,1\}$, recursively defined by the following clauses:

1. $i(x) = v(x), x \in Var$
2. $i(\neg x) = \neg i(x), x \in \mathbf{F_p}$
3. $i(x \rightarrow y) = i(x) \rightarrow i(y), x, y \in \mathbf{F_p}$

where the symbols $\neg$, $\rightarrow$ on the right hand side of the identities 2. and 3. are the propositional connectives negation and implication, semantically defined by the following truth tables:

for the unary connective $\neg$

x	$\neg$x
0	1
1	0

for the binary connective $\rightarrow$

x	y	x $\rightarrow$ y
0	0	1
0	1	1
1	0	0
1	1	1

[14] In fact, it can be proved that i is the homomorphic extension of the valuation function v over the set $\mathbf{F_p}$ of formulas.

Definition (truth):
A formula x of $\mathbf{C_p}$ is **true** under an interpretation i iff i(x) = 1

Definition (model):
A **model** of a formula x is any interpretation i such that i(x) = 1

Definition (satisfiability):
A formula x is said to be **satisfiable** iff there is an interpretation i such that i(x) = 1, or, in the terms of the above definition, iff the formula has a model.

Definition (tautology):
A formula x is said to be a **tautology** iff for every interpretation i, i(x) = 1, or, equivalently, if every interpretation is a model of the formula.
Formally: $\models x$

Definition (consequence):
A formula x is a **consequence** of a set $\Gamma = \{y_1, y_2, \ldots, y_n\}$ of formulas iff there is no interpretation i such that $i(y_1) = i(y_2) = ... = i(y_n) = 1$ and i(x) = 0, or, equivalently iff there is no model of the set Γ of formulas which is not a model of x also.

III. Kalmár's completeness

Now we can state and prove the lemma which constitutes the basis of Kalmár's completeness theorem. As we have mentioned before, this lemma sets up a correspondence between the syntactical side and the semantic side of propositional calculus, and the lemma's proof offers us an effective procedure by which we can correlate the two.

Kalmár's lemma: Let x be a formula of propositional calculus consisting only of the propositional variables $p_1, ..., p_m$ i.e $x(p_1, ..., p_m)$ and i an interpretation of the variables $p_1, ..., p_m$. We define:

(1) $p_k^i = \begin{cases} p_k, \text{if } i(p_k) = 1 \\ \neg p_k, \text{if } i(p_k) = 0 \end{cases}$, where $k \in \{1, \ldots, m\}$

and

(2) $x^i = \begin{cases} x, \text{if } i(x) = 1 \\ \neg x, \text{if } i(x) = 0 \end{cases}$

Then:

$p_1^i, p_2^i, ..., p_m^i \vdash x^i$

Demonstration: by induction on the complexity of x.

Base: c(x) = 0. According to the definition of the complexity of a formula and the unique decomposition theorem, $x = p_k$. In this case where we have $x(p_k)$, the number of possible interpretations is $2^1 = 2$:

a) $i(p_k) = 1$ or

b) $i(p_k) = 0$.

Corresponding to situation a) and definition (1) we have $p_k^i = p_k$. But $x = p_k$ and by definition (2) and situation a) we also have that $x^i = x = p_k$. According to the AS property we have:

(i) $p_k \vdash p_k$

that is: $p_k^i \vdash x^i$

Corresponding to situation b) and to definitions (1) and (2) we have $p_k^i = \neg p_k$ and $x^i = \neg x = \neg p_k$. According to the AS property we have:

(ii) $\neg p_k \vdash \neg p_k$

that is: $p_k^i \vdash x^i$

So, for c(x) = 0, the lemma holds. We now have to prove the inductive step.

Inductive step: let c(x) = n, n > 0. According to the inductive hypothesis we admit that the lemma holds for any k < n. According to the definition of the complexity and the unique decomposition theorem the formula x can have only two forms: $(\neg y)$ or $(y \rightarrow z)$, where c(y) < n and c(z) < n. We distinguish two cases, depending on the two forms of the formula x:

Case I: $x = (\neg y)$, where y and x have the same propositional variables i.e $y(p_1, ..., p_m) = x(p_1, ..., p_m)$, and c(y) < n. Hence, by induction step, the lemma holds for y^i. But depending on i(y) we have two subcases:

Subcase I a. i(y) = 0, hence $y^i = \neg y$. Knowing that $x = (\neg y)$, according to the truth table of ¬ we have i(x) = 1. So, according to point (2) of the above

definition we have: $x^i = x = (\neg y)$. Given that $c(y) < n$ we have, by the induction hypothesis:

(I.H) $p_1^i, p_2^i, ..., p_m^i \vdash y^i$

that is,

(iii) $p_1^i, p_2^i, ..., p_m^i \vdash \neg y$

From (iii) and the identity $x^i = x = (\neg y)$ results:

(α) $p_1^i, p_2^i, ..., p_m^i \vdash x^i$

Subcase I b. $i(y) = 1$, hence $y^i = y$. Knowing that $x = (\neg y)$, according to the truth table of ¬ we have that $i(x) = 0$. So, according to point (2) of the above definition we have: $x^i = (\neg x) = (\neg(\neg y))$. Given that $c(y) < n$ we have, by the induction hypothesis:

(I.H) $p_1^i, p_2^i, ..., p_m^i \vdash y^i$

that is,

(iv) $p_1^i, p_2^i, ..., p_m^i \vdash y$

from theorem 1 and THIN we have:

(v) $p_1^i, p_2^i, ..., p_m^i \vdash (y \rightarrow (\neg(\neg y)))$

From (iv) and (v):

(vi) $p_1^i, p_2^i, ..., p_m^i \vdash (\neg(\neg y))$, mp (iv), (v)

From (vi) and the identity $x^i = (\neg x) = (\neg(\neg y))$ it results that:

(β) $p_1^i, p_2^i, ..., p_m^i \vdash x^i$

Case II: $x = (y \rightarrow z)$, y and z having equal or less propositional variables than x, namely $y(p_1, ..., p_j)$, $z(p_1, ..., p_k)$, $j, k \leq m$, with $c(y) < n$ and $c(z) < n$. Hence, by induction step, the lemma holds for y^i and z^i. We distinguish three relevant cases, depending on the values of y and z, i.e depending on i(y) and i(z):

Subcase II a. $i(y) = 0$, hence $y^i = \neg y$ [z's value doesn't matter [15] – it can be $i(z) = 0$ or $i(z) = 1$]. Knowing that $x = (y \rightarrow z)$, according to the truth table of → we have $i(x) = 1$. So, according to point (2) of the above definition we

[15] if i(y) = 0, then, according to the truth table of implication, z's value is irrelevant to the construction of x^i's proof

have $x^i = x = (y \rightarrow z)$. Given that $c(y) < n$ we have, by the induction hypothesis:

(I.H) $p_1^i, p_2^i, ..., p_j^i \vdash y^i$

From THIN[16] and (I.H) we have:

(vii) $p_1^i, p_2^i, ..., p_m^i \vdash (\neg y)$

from theorem 2 and THIN we have:

(viii) $p_1^i, p_2^i, ..., p_m^i \vdash ((\neg y) \rightarrow (y \rightarrow z))$

from (vii) and (viii):

(ix) $p_1^i, p_2^i, ..., p_m^i \vdash (y \rightarrow z)$ mp (vii), (viii)

From (ix) and the identity $x^i = x = (y \rightarrow z)$ it results that:

(γ) $p_1^i, p_2^i, ..., p_m^i \vdash x^i$

Subcase II b. $i(z) = 1$, hence $z^i = z$ [y's value doesn't matter [17] – it can be $i(z) = 0$ or $i(z) = 1$]. Knowing that $x = (y \rightarrow z)$ according to the truth table of $\rightarrow$ we have $i(x) = 1$. So, according to point (2) of the above definition we have $x^i = x = (y \rightarrow z)$. Given that $c(z) < n$ we have, by the induction hypothesis:

(I.H) $p_1^i, p_2^i, ..., p_k^i \vdash z^i$

From THIN and (I.H) we have:

(x) $p_1^i, p_2^i, ..., p_m^i \vdash z$

from theorem 4 and the THIN property we have:

(xi) $p_1^i, p_2^i, ..., p_m^i \vdash (z \rightarrow (y \rightarrow z))$

from (x) and (xi):

(xii) $p_1^i, p_2^i, ..., p_m^i \vdash (y \rightarrow z)$ mp (x), (xi)

from (xii) and the identity: $x^i = x = (y \rightarrow z)$ it results that:

(δ) $p_1^i, p_2^i, ..., p_m^i \vdash x^i$

---

[16] It is possible for the formula x to have the same propositional variables as one or both of the subformulas y and z. Obviously, in this case, we needn't apply the THIN property.

[17] if $i(z) = 1$, then, according to the truth table of implication, y's value is irrelevant to the construction of x^i's proof

Subcase II c. i(y) = 1 and i(z) = 0, hence $y^i = y$ and $z^i = \neg z$. Knowing that x = (y → z), according to the truth table of →, we have i(x) = 0. So, according to point (2) of the above definition, we have $x^i = \neg x = \neg(y \rightarrow z)$. Given that c(y) < n and c(z) < n, by the induction hypothesis:

(I.H) $p_1^i, p_2^i, ..., p_j^i \vdash y^i$

(I.H) $p_1^i, p_2^i, ..., p_k^i \vdash z^i$

From THIN and (I.H) we have:

(xiii) $p_1^i, p_2^i, ..., p_m^i \vdash y$

(xiv) $p_1^i, p_2^i, ..., p_m^i \vdash \neg z$

from theorem 5 and the THIN property we have:

(xv) $p_1^i, p_2^i, ..., p_m^i \vdash (y \rightarrow ((\neg z) \rightarrow (\neg(y \rightarrow y))))$

from (xiii) and (xv):

(xvi) $p_1^i, p_2^i, ..., p_m^i \vdash ((\neg z) \rightarrow (\neg(y \rightarrow z)))$ mp (xiii), (xv)

and from (xiv) and (xvi):

(xv) $p_1^i, p_2^i, ..., p_m^i \vdash \neg(y \rightarrow z)$ mp (xiv), (xvi)

from (xv) and the identity $x^i = \neg x = \neg(y \rightarrow z)$ it results that:

(ε) $p_1^i, p_2^i, ..., p_m^i \vdash x^i$

From (α), (β), (γ), (δ), (ε) we get the proof for the induction step and with this the proof of Kalmár's lemma.

Now we can prove:

Completeness Theorem: If $\models x$ then $\vdash x$

Demonstration: Let x be any formula of the propositional calculus consisting of the variables $p_1 - p_m$ i.e $x(p_1, ..., p_m) \in \mathbf{F_p}$. From the theorem's hypothesis we have $\models x$, that is, according to the definition of tautology, i(x) = 1and hence $x^i = x$, for any interpretation of propositional variables $i(p_1, ..., p_m)$. We define the two interpretations i and j of the propositional variables which occur in the formula x as:

a) $i(p_1) = i(p_2) = ...i(p_{m-1}) = 1, i(p_m) = 1$

b) $j(p_1) = j(p_2) = ...i(p_{m-1}) = 1, i(p_m) = 0$

According to the definition of the interpretations i and j, and to the tautological character of the formula x which gives us the identity: $x^i = x^j = x$, Kalmár's lemma allows us to assert that:

a) $p_1^i, p_2^i, ..., p_m \vdash x$

b) $p_1^j, p_2^j, ..., \neg p_m \vdash x$

By applying the deduction theorem to a) and b) we get:

1. $p_1^i, p_2^i, ... p_{m-1}^i \vdash p_m \rightarrow x$

2. $p_1^j, p_2^j, ..., p_{m-1}^j \vdash \neg p_m \rightarrow x$

Because i and j assign the same values to $p_1, \ldots, p_{m-1}$, $p_k^i = p_k^j$, for all $k < m$.

From theorem 6 and the THIN property we have:

3. $p_1^i, p_2^i, ... p_{m-1}^i \vdash ((p_m \rightarrow x) \rightarrow (((\neg p_m) \rightarrow x) \rightarrow x)$

4. $p_1^i, p_2^i, ... p_{m-1}^i \vdash (((\neg p_m) \rightarrow x) \rightarrow x)$ mp 1, 3

5. $p_1^i, p_2^i, ... p_{m-1}^i \vdash x$ mp 2, 4

By repeating the above given algorithm m times we will obtain:

$\vdash x$,

in other words, the conclusion of the theorem. Once we reached this final step, the theorem is proved.

IV. How effective is Kalmár's lemma, anyway?

The object of our debate, though, will be Kalmár's lemma. Two important aspects of the proof should be highlighted: the first is that it has an effective character and the second is that it uses the mathematical induction method in its proof. Without getting into a detailed discussion about the properties of inductive sets, let us begin with an observation regarding the way in which we can consider the proof's effective character.

The proof of Kalmár's lemma specifically states how this connection is accomplished: by mirroring semantical computations with syntactical ones. Basically, the lemma constructs, by definitions (1) and (2), the syntactical counterparts of each truth value assigned to propositional variables and formulas by a particular interpretation and then its proof

provides a way of computing deducibility relations between these syntactical counterparts. For example, let's start with a particular interpretation of the propositional variables of an arbitrary formula x. According to "the central thesis of propositional logic"[18], any assignment of truth values to the propositional variables of a formula x, "can be extended, by means of the truth-table definitions […], to give a truth-value to [x]; this truth-value assigned to x is uniquely determined and it can be computed mechanically"[19].

So, given a particular interpretation like the one abovementioned, we can semantically compute the truth value of a formula x. This computation is performed, as has been said, by means of truth tables. The proof of Kalmár's lemma shows how to mimic these semantical computations by syntactical ones. By definition (1) we establish the syntactical counterparts of the truth value assigned, by this particular interpretation, to the propositional variables of the formula x. These syntactical counterparts and the AS property of the deducibility relation form the base case of the lemma. Next, we employ one of the theorems 1 – 4 according to each step which mimics the semantical computations.

Finally what was obtained by semantic computation, that is the truth value of the formula x in that particular interpretation, is mirrored by the construction of a deducibility relationship between the syntactical counterparts of the truth value of the propositional variables of the formula x and the syntactical counterpart of the truth value of the formula x. This is the reason why we can say that the lemma provides the syntactical means by which we can mirror the semantic computations done by the method of truth tables. In other words, for each row of the truth table the lemma provides a corresponding deducibility relation.

A perspective, which we will call bottom-up, sees the proof of the lemma as a way of constructing the deducibility relationship between the formula x^i and its propositional variables, p_1^i, p_2^i, ..., p_m^i, that begins with a

[18] Adequately named like this by Wilfrid Hodges in Dov Gabbay, Franz Guenthner [15] , p. 11

[19] Dov Gabbay, Franz Guenthner [15], p. 11

particular valuation of its propositional variables and its syntactical counterparts as specified by definition (1) in Kalmár's lemma. The construction is "from bottom to top" because it starts with the truth values assigned to each propositional variables by that particular valuation, and their syntactical correspondent, which forms the basis of the lemma, and it moves up to the formula x^i, by applying one of the theorems 1 – 4 to each new level of construction.

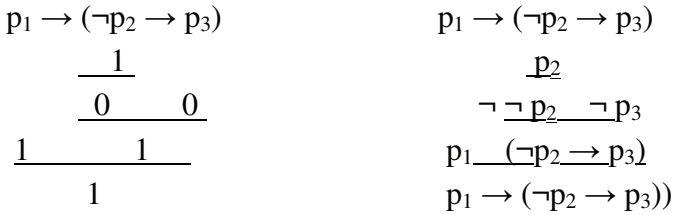
Basically, we begin with the base case of the lemma provided by the syntactical correspondent of an arbitrary interpretation of a formula's propositional variables, construct a deducibility relation, in an ascending order, to each subformulas of the formula, according to the syntactical correspondent of their truth value in this arbitrary valuation and move up to the point where we reach the deducibility relation of the truth value of the formula itself under this interpretation. One of the representatives of this perspective seems[20] to be Stephen Cole Kleene.

In order to clarify the way in which we can determine the "effectiveness" of the proof of Kalmár's lemma from the bottom - up perspective, let us take an adequate example for the system we have built so far.

Let x be the following formula:
$x = (p_1 \rightarrow (\neg p_2 \rightarrow p_3))$, and consider the following interpretation:
$i(p_1) = 1$
$i(p_2) = 1$
$i(p_3) = 0$.

Then the bottom - up construction of the deducibility relation between x^i and its propositional variables can be viewed by the means of a diagram composed of two trees corresponding to the two types of computation: semantic in the left tree and syntactical in the right tree.

[20] At least this is how I read his commentaries and examples on Kalmár's lemma, in both his *Introduction to metamathematics* and *Mathematical logic*; for further details see Stephen Cole Kleene [16] § 12 and Stephen Cole Kleene [17] § 29.



This is what Kleene says concerning a similar example which illustrates the correspondence between the semantic computations given by means of the truth tables and their syntactical counterparts:

> It may be instructive to view in a diagram with two "trees" how each computation step (horizontal line in the left tree) corresponds to a deducibility relationship of [Kalmár's lemma] (horizontal line in the right tree). […] it follows that in the right tree each formula is deducible from the distinct formulas occurring at the tops of branches over it (or any larger set of formulas)[21].

Let us explore, with the aid of this example, a little more this correspondence between the two computations:

1. for $i(p_2) = 1$ [in the diagram, this is the first line below the formula] we get, by semantic computations, $i(\neg p_2) = 0$ [in the diagram, this is the second line below the formula]; correspondingly, according to definition (1) we have: $p_2^i = p_2$ and by syntactical computation from p_2 we get $\neg\neg p_2$, that is $p_2 \vdash \neg\neg p_2$. If, in the case of semantic computation, we get $i(\neg p_2) = 0$ from $i(p_2) = 1$, by means of the truth table for $\neg$, in the case of syntactical computation we get $\neg \neg p_2$ from p_2 by means of theorem 1 (with $x = p_2$)

2. for $i(\neg p_2) = 0$ and $i(p_3) = 0$ [in the diagram, this is the second line, below the formula] we get, by semantic computations, $i(\neg p_2 \rightarrow p_3) = 1$ [in the diagram, this is the third line below the formula]; correspondingly,

[21] Stephen Cole Kleene [16], p. 47

according to definition (1) and (2) we have $\neg \neg p_2$, $\neg p_3$, $(\neg p_2 \rightarrow p_3)$ and by syntactical computation, from $\neg \neg p_2$, $\neg p_3$ we get $(\neg p_2 \rightarrow p_3)$, that is $\neg \neg p_2$, $\neg p_3 \vdash (\neg p_2 \rightarrow p_3)$. If, in the case of semantic computation, we get $i(\neg p_2 \rightarrow p_3) = 1$ from $i(\neg p_2) = 0$ and $i(p_3) = 0$, by means of the truth tables for $\rightarrow$, in the case of syntactical computation we get $(\neg p_2 \rightarrow p_3)$ from $\neg \neg p_2$ and $\neg p_3$ by means of theorem 2 (with $x = \neg p_2$ and $y = p_3$).

By repeating this process we will finish by constructing a deducibility relation between the formula x^i and its propositional variables $p_1^i, p_2^i, ..., p_m^i$, under this particular interpretation, of course, which is what Kalmár's lemma had set out to do.

This perspective on Kalmár's lemma about the inductive construction of the deducibility relationship starting from the valuations of propositional variables and moving up to the formula itself can be reversed.

More exactly, starting from a given formula $x(p_1, ..., p_m)$ of the propositional calculus, Kalmár's lemma states the existence of a deducibility relationship between the formula x^i and its propositional variables $p_1^i, p_2^i, ..., p_m^i$, for any interpretation of the formula x. Given a truth value assigned to the formula x by a particular interpretation, we can, by semantic computations done by means of truth tables, determine the truth values of its immediate subformulas. By repeating this procedure we will get the truth value of its propositional variables. As we have argued when we discussed the details of Kalmár's lemma, these semantic computations are mirrored by syntactical ones. So, by the syntactical counterpart of the abovementioned procedure we will effectively reduce the deducibility relationship between the formula x^i and its propositional variables $p_1^i, p_2^i, ..., p_m^i$ to the deducibility relationship between the immediate subformulas and their propositional variables to the last of the formula's x^i fundamental assumptions, that is to those assumptions consisting only of propositional variables, which, obviously, form the base case of the lemma. The decomposition of the deducibility relationship of the formula in question into smaller deducibility relationships exploits the formulas' syntactical

structure, namely that we can determine the immediate subformula/subformulas of any formula that has a complexity > 0. This perspective, let us call it "top – down", seems to have been endorsed by Alonzo Church when he writes[22]:

> The proof of [Kalmár's lemma] is effective in the sense that it provides an effective method for finding a proof of $[x^i]$ from the hypotheses $[p_1^i, p_2^i, ..., p_m^i]$. If x has no occurrences of →, this is provided directly. If $[x^i]$ has occurrences of →, the proof provides directly an effective reduction of the problem of finding a proof of $[x^i]$ from the hypotheses $[p_1^i, p_2^i, ..., p_m^i]$ to the two problems of finding proofs of y^i and z^i $[x = (y \rightarrow z)]$ from the hypotheses $[p_1^i, p_2^i, ..., p_m^i]$; the same reduction may then be repeated upon the two latter problems, and so on; after a finite number of repetitions the process of reduction must terminate, yielding effectively a proof of $[x^i]$ from the hypotheses $[p_1^i, p_2^i, ..., p_m^i]$[23].

In what follows I want to pinpoint a small problem with this effective "reduction" and then try to show how it can be "solved". In order to better understand what this problem is, let us describe this "top – down" perspective with the aid of an example: suppose we have an arbitrary formula x of $\mathbf{C_p}$, $x = (p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)$, where $x(p_1, p_2)$. Therefore the lemma asserts:

(K.L) $p_1^i, p_2^i \vdash ((p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1))^i$

According to the "top – down" perspective, the inductive proof of the lemma allows us to reduce this deducibility relationship to:

$p_1^i, p_2^i \vdash (p_1 \rightarrow p_2)^i$

or/and

---

[22] in his book *Introduction to Mathematical Logic*, Alonzo Curch [13] uses a system of propositional calculus with just one primitive connective, namely implication (→) and a primitive constant (*f*) falshood; this is the reason why the formulas of the calculus, other than variables and constant, contain only the implication connective. Negation, for example, is defined using implication and falshood in the following way: $\neg p = (p \rightarrow f)$.

[23] Alonzo Curch [13], pp. 98-99

$p_1^i, p_2^i \vdash (\neg p_2 \rightarrow \neg p_1)^i$

which, by repeated reductions, results in:

$p_1^i, \vdash p_1^i$

$p_2^i \vdash p_2^i$

Let's consider an interpretation i in which the formula x is true, i(x) = 1.

Then $x^i = ((p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1))^i = (p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)$

According to Kalmár's lemma:

(K.L) $p_1^i, p_2^i \; p_3^i \vdash x^i$ that is:

$p_1^i, p_2^i \vdash (p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)$

Let us try, in conformity with the "top – down" perspective, to decompose this deducibility relationship into simpler deducibility relationships[24]. Given the fact that x has the form (y → z), where y = (p1 → p2) and z = (¬p2 → ¬p1) and that i(x) = 1, by the truth table for →, we have either i(y) = 0, hence $y^i = \neg y = \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$ or i(z) = 1, hence $z^i = (\neg p_2 \rightarrow \neg p_1)$. If i(y) = 0 then we find ourselves under subcase **IIa**. If i(z) = 1 then we find ourselves under subcase **IIb** of the lemma. These subcases allow us, given that c(y) < c(x) and c(z) < c(x), to assert:

$p_1^i, p_2^i \vdash \neg y$ [according to the subcase **IIa**]

or

$p_1^i, p_2^i \vdash z$ [according to the subcase **IIb**]

Knowing that y = (p1 → p2) and z = (¬p2 → ¬p1) the above relations are:

$p_1^i, p_2^i \vdash \neg(p_1 \rightarrow p_2)$ [using theorem 2]

or

$p_1^i, p_2^i \vdash (\neg p_2 \rightarrow \neg p_1)$ [using theorem 3]

In this way we effectively reduced the deducibility relation between the propositional variables p_1^i, p_2^i of the formula x, and the formula x^i, under this particular interpretation, to the deducibility relation between the propositional variables p_1^i, p_2^i of the immediate subformulas y and z, and the immediate subformulas y^i and z^i

---

[24] simple, here, means with a smaller complexity

The legitimacy of the first reduction is assured by theorem 2, and the legitimacy of the second by theorem 3, more precisely if we have

1. $p_1^i, p_2^i \vdash \neg(p_1 \rightarrow p_2)$

and we use theorem 3 with $x = (p_1 \rightarrow p_2)$ and $y = (\neg p_2 \rightarrow \neg p_1)$ then we get:

2. $p_1^i, p_2^i \vdash ((\neg(p_1 \rightarrow p_2)) \rightarrow ((p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)))$

From 1. and 2. we obtain:

3. $p_1^i, p_2^i \vdash (p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)$ mp, 1. 2.

Or if we have:

4. $p_1^i, p_2^i \vdash (\neg p_2 \rightarrow \neg p_1)$

and we use theorem 4 with $x = (p_1 \rightarrow p_2)$ and $y = (\neg p_2 \rightarrow \neg p_1)$ then we get:

5. $p_1^i, p_2^i \vdash ((\neg p_2 \rightarrow \neg p_1) \rightarrow ((p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)))$

From 4. and 5. we obtain:

6. $p_1^i, p_2^i \vdash (p_1 \rightarrow p_2) \rightarrow (\neg p_2 \rightarrow \neg p_1)$ mp, 4. 5.

As we have mentioned above when we described the “top – down” perspective, by systematically repeating the above given procedure we will come to the deducibility relationship between the propositional variables and their negations, that is to $p_k \vdash p_k$ or $\neg p_k \vdash \neg p_k$. For instance, if we take into consideration alternative 4. from the above given example, then the deducibility of the formula x^i is reduced to the deducibility of 4. i.e

$p_1^i, p_2^i \vdash (\neg p_2 \rightarrow \neg p_1)$

which, in turn, decomposes into other two deducibility relationships, namely:

7. $p_1^i, p_2^i \vdash p_2$ [using theorem 3] or

8. $p_1^i, p_2^i \vdash \neg p_1$ [using theorem 4]

At this stage we have reached the base case of Kalmár’s lemma So far so good. But let us now take into consideration the following formula x of $\mathbf{C_p}$: $x = \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$ and see how it can be reduced, and what lemmas we can use to reduce it. Obviously, x has only three propositional variables, p_1, p_2, p_3 that is $x(p_1, p_2, p_3)$

We assume that the formula is true, that means there is an interpretation i so that $i(x) = 1$. Therefore, we have:

$x^i = \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$. According to Kalmár's lemma:

(K.L) $p_1^i, p_2^i\ p_3^i \vdash x^i$, that is

$p_1^i, p_2^i\ p_3^i \vdash \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$

Accordingly to the "top – down" perspective, we will try to decompose this deducibility relationship into simpler deducibility relationships, as we have done in the previous example. Given the fact that $x = (\neg y)$, where $y = (p_1 \rightarrow (p_2 \rightarrow p_3))$ and $i(x) = 1$ we have, by the truth table for $\neg$, $i(y) = 0$, hence $y^i = \neg y = \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$. What is remarkable in this particular case is that, according to definitions (1) and (2) from Kalmár's lemma, x^i and y^i have the same form although $c(y) < c(x)$, y being an immediate subformula of x. In this situation, we find ourselves under subcase **Ia** of the lemma, which allows us, given that $c(y) < c(x)$, to assert:

$p_1^i, p_2^i\ p_3^i \vdash y^i$

But $y^i = \neg y$, so:

$p_1^i, p_2^i\ p_3^i \vdash \neg(p_1 \rightarrow (p_2 \rightarrow p_3))$

which is our formula x. Apparently by applying the method which Kalmár used in his proof of the lemma to decompose the deducibility relationship between this formula x^i and its propositional variables p_1^i, p_2^i p_3^i into simpler relationships we reach to the same deducibility relation. Therefore, in this case of the lemma, the decomposition is not as "effective" as Alonzo Church thought. Why does our attempt to decompose the deducibility relationship into simpler deducibility relationships fail in the subcase **Ia** of the lemma? The answer is that, in this subcase, the proof of the lemma is assured by the identity of x^i and y^i although it is clear that y is an immediate subformula of x. In other words, in this subcase we do not actually decompose the deducibility relationship of the formula x^i into the deducibility relationship of the subformula y^i, but we take advantage of the fact that the syntactical form corresponding to the interpretation of the formula x is identical with the syntactical form corresponding to the subformula y, that is $x^i = y^i$. As we can see from the lemma's proof, this is the only case which does not require the use of any helping theorem,

because the step from the formula x^i to its subformula y^i is an immediate one.

Well, all is not lost. This situation can be rectified if we proceed to the following subdividing of case **I**, according to the complexity of the formula y:

Case I'

Ia'. x = (¬y), i(y) = 0 and c(y) = 0

Then the solving procedure is reduced to the basis of lemma and the AS property.

Ia''. x = (¬y), i(y) = 0 and c(y) > 0

Then y has either the form (z → t) or the form (¬z).

1. y = (z → t). Knowing that i(y) = 0 we have i(z → t) = 0; by the truth table for → we have i(z) = 1 and i(t) = 0, that is we find ourselves under case **IIc** of Kalmár's lemma with y = z and z = t ; consequently we treat this case accordingly.
2. y = (¬z). Knowing that i(y) = 0 we have i(¬z) = 0 by the truth table for ¬ we have i(z) = 1. This is the case **Ib** of Kalmár's lemma with y = z; consequently we treat this case accordingly.

Obviously, the previous considerations do not imply a flaw of the proof, they only highlight two ways of understanding Kalmár's proof of the lemma and certain matters these two perspectives bring up regarding the lemma's effective character, which the author did not find treated as such anywhere in the specialized literature he consulted. Insofar as we accept these perspectives on the lemma, and we classify different authors according to them, then the effectiveness of the lemma's character becomes, to a certain extent, a problematical matter, at least as far as the "top – down" perspective is concerned. I say to a certain extent because, although the lemma's proof, as it is, doesn't entirely have an effective character it can nevertheless be rectified, following the above mentioned suggestion.

References:

[1] Kurt Gödel [1930], "Die Vollständigkeit der Axiome des logischen Funktionenkalküls" in *Monatshefte für Mathematik und Physik* 31, 349-360

[2] Leon Henkin [1949a], "The completeness of the first-order functional calculus" in *The Journal of Symbolic Logic* 14: 159-166

[3] Jaakko Hintikka [1955], "Form and content in quantification theory", *Acta Philosophica Fennica*, 8: 11-55

[4] Ian Chiswell and Wilfrid Hodges [2007], *Mathematical Logic*, Oxford: Oxford University Press

[5] Evert W. Beth [1953], "A topological proof of the theorem of Löwenheim-Skolem-Gödel", *Indagationes Mathematicae*, 15: 66-71

[6] Jerzy Łos [1951], "An algebraic proof of completeness for the two valued propositional calculus*", Colloquium Mathematicum*, 2: 236-240

[7] Helena Rasiowa, Roman Sikorski [1963], *The mathematics of metamathematics*, Warszawa: Panstwowe Wydawnictwo Naukowe

[8] Marshall H. Stone [1936], "The theory of representations for Boolean algebras", *Transactions of the American Mathematical Society*, 40: 37–111

[9] John W. Dawson [1993], "The compactness of first-order logic: from Gödel to Lindström", *History and Philosophy of Logic*, 14: 15-37

[10] Emil Post [1921], "Introduction to a general theory of elementary propositions" in *American Journal of Mathematics* 43: 163-185

[11] Bertrand Russell, Alfred Whitehead [1910], *Principia Mathematica*, Cambridge, UK: Cambridge University Press

[12] László Kalmár [1935], "Über die Axiornatisierbarkeit des Aussagenkalküls", *Acta Scientiarum Mathematicarum* 7: 222-243

[13] Alonzo Church [1956], *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press

[14] Leon Henkin [1949b], "Fragments of the propositional calculus", *Journal of Symbolic Logic*, 14: 42-48

[15] Dov Gabbay, Franz Guenthner [1983], *Handbook of philosophical logic*, Dordrecht, Boston, Lancaster: D. Reidl

[16] Stephen Cole Kleene [2002] *Mathematical logic*, Mineola, New York: Dover

[17] Stephen Cole Kleene [1952], *Introduction to metamathematics*, Amsterdam: North-Holland

Compatibilism vs. Incompatibilism: An Integrated Approach from Participant Stance and Affect

Sharmistha DHAR [*]
Department of Philosophy
Centre for Cognitive Science
Jadavpur University Kolkata, India

Abstract:

Following the recent surge in experimental philosophy exploring how unprimed intuitions enable the folk arrive at judgments concerning free will and moral responsibility, a widespread anomaly in folk intuitions has been reported. This has given rise to two different explanatory frameworks- one counting on affect that has been projected as making all the difference between compatibilism and incompatibilism and the other relying on Strawsonian participant attitude while accounting for compatibilist responses. The aim of this paper is to bring to the fore the asymmetric folk intuitions regarding ascription of moral responsibility, the expository accounts- one put forward by Shaun Nichols and the other by Eddy Nahmias, and show possibility of reconciliation between the two apparently different views, especially when it comes to unravelling the psychological mechanism underlying compatibilist intuition.

Keywords: Compatibilism, Incompatibilism, Affect, Participant Stance, Mechanistic Stance.

1. Introduction

Among the philosophical fraternity, there seems to be no unanimity regarding whether it is compatibilist intuition or incompatibilist intuition that should be given due weightage. On the one hand, there are staunch

[*] E-mail: sharmistha.dhar@rediffmail.com.

incompatibilists like Galen Strawson and Laura Ekstrom who are convinced that it is "in our nature to take determinism to pose a serious problem for our notions of responsibility and freedom"[1] and that "we come to the table, nearly all of us, as pretheoretic incompatibilists".[2] There are philosophers like Daniel Dennett, on the other end of the spectrum who claim that ordinary people care two hoots about whether a convict could have done otherwise while trying to determine whether the person is to be exonerated or proclaimed guilty [3] - they have a natural compatibilist orientation. Susan Wolf notes that compatibilism "seems to accord with and account for the whole set of our intuitions about responsibility better than … the leading alternatives " [4] While plumbing the literature that makes this debate its centrepiece, what we find is a uniform appeal by philosophers on both sides to draw on pre-theoretic, folksy intuitions. This accounts for the shift of attention to the descriptive question of what are the natural responses of the laypeople while ascribing moral accountability and what makes them judge what they do. We will then scan through the experimental results in the next section and discuss the most viable of psychological mechanisms underlying these intuitions in the third section.

2. Mapping how Folk Intuitions Shape Judgments of Moral Responsibility

Shaun Nichols and his associate Joshua Knobe ran an experiment to ascertain whether participants envisage human behaviour, especially choices and decisions as deterministic (causally inevitable) or indeterministic (contingent upon the agent's belief and desire). They presented the participants with the description of two universes. Now, most participants chose Universe B (in which choices and decisions *did not have to happen the way they did*, by virtue of the fact that antecedent conditions were incapable of "calling the shots") over Universe A (in which choices and

[1] Strawson: 1986, p.89.
[2] Ekstrom: 2002, p.310.
[3] Dennett: 1984, p.558.
[4] Wolf: 1990, p. 89.

decisions *had to happen the way they did*, every choice and decision being completely caused by their antecedent conditions).

Let us briefly describe the test conditions. Participants were given the impression that in Universe A, the domain of human behaviour is such that, the coming into being of any choice or decision is the reflection of a rule or a law that the prior conditions of that particular choice always make its occurrence necessary and irreversible. Universe B, in sharp contrast, was designed not to come under such a rule insofar as the domain of human behaviour was concerned.[*] Just as the Universe A condition could lead the participants to believe in the logical possibility (if not empirical) of predicting an agent's act by dint of knowledge of its antecedent conditions, the Universe B condition also gave reason to believe in the empirical possibility that at least human choice-making events could be spared from any causal necessity (such a possibility was stoked by the phrase: "…even if everything in the universe was exactly the same up until Mary made her decision, it *did not have to happen* that Mary would decide to have French Fries"). And when the time came for them to identify which of these two beliefs they found more reliable than the other, we know that an overwhelming number of participants sided with Universe B (the indeterministic universe). Although Nichols' purpose to ask this initial question was "simply to see whether subjects believe that our own universe is deterministic or indeterministic", we may take this result as an indicator of two vying possibilities:

1) The folk are staunch indeterminists, inveterately agent-causationist style; they may of course be libertarian indeterminists without being agent-causationists. They gauge an agent's freedom of action and will by considering whether the person in question caused that action by dint of his own will; and that being the case, they believe that it is quite an

[*] The description of Universe B read: "Now imagine a universe (Universe B) in which *almost* everything that happens is completely caused by whatever happened before it. The one exception is human decision-making" (Nichols: 2007, p. 673).

(empirical) possibility that Mary could have chosen to have something other than French Fries as she, like all humans could but be left on her own will. They think that such a possibility will be marred by a *necessitarian* causal law. The folk are thus incompatibilists.

2) The folk believe that it is important that an agent is able to do otherwise than he originally wanted to. However, this ability to exercise a climb-down is made possible only when the agent modifies his original belief states or desire states or plans. Mary could have had an ice cream instead of French Fries only if she wanted to (a change in her desire state ensured it). The folk may think that it is so obvious that one needs not even make a mention of it. The folk thus might be psychological determinists and still compatibilists.

Now following the empirical work of Nichols and Knobe on folk intuitions, we will try to find out which of these two possibilities gains more credence.

2.1. Nichols' and Knobe's Findings [5]

a) In this experiment, immediately following the Universe task, participants were randomly assigned either to the *abstract* condition or to the *concrete* condition. The *concrete* scenario read:

In Universe A, a person named Bill murders his wife and children by detonating an explosive at his home with the single motive of being with his secretary with whom he has developed an illicit relationship.

Participants were then presented with the question: Is Bill fully morally responsible for killing his wife and children?

YES NO

The participants in the *abstract* condition instead received just the question, which was however couched in a fashion that prompted them to think in a more general way. The question posed to them was: In Universe

[5] For details of this empirical research, see Nichols: 2007a and Nichols (forthcoming) respectively.

A is it possible for a person to be fully morally responsible for his or her actions?

YES NO

Table 1 shows the results:

	Compatibilist Responses	**Incompatibilist Responses**
Concrete Condition	72%	NA
Abstract Condition	NA	86%

Table 1

Nichols and Knobe, however, were a bit skeptical whether the "prolixity" of the *concrete* condition took its toll on the subjects who as a result, forgot that the heinous crime was perpetrated in a deterministic Universe. They, therefore, ran the *concrete* condition once more, making the condition a little terse. It now read:

In Universe A, Bill stabs his wife and children to death so that he can be with his secretary. Is it possible that Bill is fully morally responsible for killing his family?

YES NO

And although the table was a little turned both literally and figuratively, the volume of compatibilist responses (in terms of percentage) in the *concrete* scenario was still much lower than that in the *abstract* scenario (vide Table 2).

	Compatibilist Responses	**Incompatibilist Responses**
Concrete Condition	50%	NA
Abstract Condition	NA	86%

Table 2

b) The previous experiment primarily looked into the effect of *abstract-concrete* conditions on folksy moral judgments. The results also

contained an indication that some emotive attitudes (anger, sympathy etc. in Bill's case) might spur compatibilist responses. Thus, the hypothesis Nichols wanted to test in the next experiment was whether affect engenders compatibilist intuitions. Here, Nichols once again used the concrete condition as indication has already been found that compatibilist tendencies are tied to a concrete description of a morally salient situation. He thus used the concrete condition as an independent variable and used affect as a dependent variable varying its nuances.

Participants were accordingly randomly assigned to either a *high affect* condition or a *low affect* condition. In both the conditions, half of the participants were asked to consider Universe A as the locus of the agent and his act and the other half were asked to consider Universe B where the agent lived. The descriptions of both the conditions were as follows:

High Affect Condition: As he has done many times in the past, Bill stalks and ravishes a stranger. Is it possible for Bill to be fully morally responsible for this act?

Low Affect Condition: As he has done many times in the past, Mark decides once again to dodge his taxes. Is it possible for Mark to be fully morally responsible for this act?

The results (vide Table 3 & Table 4) indicated that the influence of affect on compatibilist responses cannot be overlooked.

	High Affect Case	**Low Affect Case**
	The physical tormentor's case (Indeterministic Universe)	The tax dodger's case (Indeterministic Universe)
Percentage of Participants assigning MR	95%	89%

Table 3

	High Affect Case	Low Affect Case
	The physical tormentor's case (Deterministic Universe)	The tax dodger's case (Deterministic Universe)
Percentage of Participants assigning MR	64%	23%

Table 4

In Table 3, responses are, as expected, more compatibilist than incompatibilist, especially when the agent Bill is taken to be in the indeterministic Universe. But the emotion-steeped conditions seemed to have further provoked the subject to judge that Bill is fully morally responsible which is evident by the figure "95%" of the High Affect Condition as against the "89%" of the Low Affect Condition. In Table 4, that presents the responses of the deterministic world scenario, even the concrete condition cannot substantially evoke Judgments of MR although it did in an earlier experiment (see Table 1), when it is tempered with a low emotional content as is evidenced by the figure "23%". In sharp contrast with this response is the figure "64%" elicited by the emotion-laden condition.

It is to be noted that Nichols and Knobe ran the two previous experiments to gather evidence regarding a rampant suspicion that affect infuses an infelicity (a bias, to be precise) in folk theories and judgments of moral responsibility that then goes to trigger compatibilist intuitions. Results of the first experiment hinted that the affect-inducing concrete cases might elicit compatibilist responses while an affect-neutral abstract condition that induces us to think in a cold, cognitive way might be responsible for incompatibilist responses. However, the second experiment, according to Nichols, went a step further in projecting another pointer, that it may be not so much a difference between abstract/concrete conditions as it is between affect-neutral and affect-laden conditions that delimit

compatibilist intuitions from incompatibilist intuitions. The case in point is a curiously "low-key" performance by compatibilist intuitions on the tax cheater's case which is a concrete case all right but significantly marked by affect- neutrality.

c) Nichols tested yet another hypothesis laid down in the real/actual world versus alternate/hypothetical world. His hunch was if determinism is ensconced in an actual universe it would elicit mostly compatibilist responses; the alternate universe condition, on the other hand, would, by and large give a leeway for denial of free will and MR.

Subjects were randomly assigned either to the *actual* world condition or to the *alternate* world condition. Both the worlds were characterized by a deterministic description. Here determinism was couched in terms of genetic make-up and environmental influence. Thus it was stipulated in both the conditions that given that each decision in this world (*actual* or *alternate*) *has to happen the way it does*, any individual having the same genetic make-up and environmental influence would decide to embark on the same action because every decision is an invariable result of the past conditions- the past conditions here being genetic make-up and environmental influence. Subjects were then presented with three statements aimed at finding out how the subjects assess the relation between the deterministic condition of the world given to them and the possibility of free will and MR. They were asked to respond with various levels of agreement and disagreement.[6] We will here focus only on the MR scenario. The results are presented in Table 5 and 6.

[6] The level of agreement or disagreement was based on a scale of 1 to 7 where 1 corresponded to complete disagreement, 4 corresponded to a neutral stance and the rating of 7 meant complete agreement. The numbers quoted in Table 5 and Table 6 indicate mean responses.

It is impossible for a person to be fully morally responsible	
Alternate Condition	**Actual Condition**
Greatly agree (5.06)	Greatly disagree (3.58)
Incompatibilist response	Compatibilist response

Table 5

People should still be morally blamed for committing crimes	
Alternate Condition	**Actual Condition**
Greatly disagree (3.67)	Greatly agree (5.35)
Incompatibilist response	Compatibilist response

Table 6

2.2. Nahmias' Findings [7]

Nahmias' experiments also produced varied responses on the question of the feasibility of moral responsibility under the shadow of determinism. He however has a different set of explanations for the emergence of the pattern of intuitions he encountered. We will first present the compatibilist responses produced by his version of a similar line of experiments cited in the foregoing.

a) Participants in this experiment were once again presented with the description of physical law determinism in the dressing of a "prophetic"

[7] For details regarding Nahmias' empirical work on folk intuitions, vide Nahmias: 2005, 2006 and 2007 respectively.

supercomputer. The deterministic proviso was couched in the following manner:

A supercomputer with the knowledge of all the laws of nature and the present state of affairs of everything in the world at its disposal can predict any future event. Thus, at a specified time, say, on March 25, 2150 AD, the supercomputer predicts that 20 years later, on January 26, 2195 AD, a person called Jeremy will rob Fidelity Bank at 6 P.M. The question put to them was whether Jeremy would be morally responsible for his misdeed. They were also asked to judge the moral responsibility of Jeremy if the supercomputer prophesied at the same manner that the Jeremy would save a child. But there was a clear majority of vote supporting that MR on that condition would not be a utopian dream (vide Table 7).

But then Nahmias did not rule out the possibility that the dose of determinism had not been strong enough while making every effort to present the concept avoiding a petitio principii. He concedes that as a result, participants were perhaps "more focused on the fact that Jeremy's actions were predicted by the supercomputer than the fact that the prediction was made based on deterministic laws". Although, he thinks "it would still be an important result that most people do not judge such *predictability* to conflict with free will and responsibility".

Is Jeremy morally responsible for his acts?	
Robbing a bank	Yes- 83%
Saving a child	Yes- 88%

Table 7

b) In the next experiment, therefore, participants were presented with an explicit description of determinism. The scenario read:

Fred and Barney are two identical twins living in a world where the beliefs and values of every person are *caused completely by the combination of one's genes and one's environment*. Now one day their mother put them

for adoption. Fred is adopted by the Jerksons and Barney is adopted by the Kindersons.

In Fred's case, his genes and his upbringing by the selfish Jerkson family have caused him to value money above all else and to believe it is OK to acquire money however you can. In Barney's case, his (identical) genes and his upbringing by the kindly Kinderson family have caused him to value honesty above all else and to believe one should always respect others' property. Both Fred and Barney are intelligent individuals who are capable of deliberating about what they do.

One day Fred and Barney each happen to find a wallet containing $1000 and the identity of the owner (neither man knows the owner). Each man is sure there is nobody else around. After deliberation, Fred Jerkson, because of his beliefs and values, keeps the money. After deliberation, Barney Kinderson, because of his beliefs and values, returns the wallet to its owner. Given that, in this world, one's genes and environment completely cause one's beliefs and values, it is true that if Fred had been adopted by the Kindersons, he would have had the beliefs and values that would have caused him to return the wallet; and if Barney had been adopted by the Jerksons, he would have had the beliefs and values that would have caused him to keep the wallet.

Once again there were more participants expressing the belief that it would be the agent himself who would be responsible for what they did despite their genetic makeup and upbringing over which they had no control (vide Table 8). That is, the responses of the majority of the participants in this experiment also bordered upon compatibilism.

Is Fred morally responsible?	**Is Barney morally responsible?**
Yes- 60%	Yes- 64%

Table 8

c) However, Nahmias et al found pre-eminently incompatibilist responses too. Just as Nichols and Knobe exposed their subjects to abstract/

concrete or real/ alternate scenarios that ended up in anomalous pattern of responses, Nahmias hoped to witness the same kind of responses by exposing his subjects to Neuro-reductionistic world versus Psychological-deterministic world scenarios. He also varied these two scenarios on the dimension of alternate world /real world conditions together with the concrete/abstract primer. That is participants were asked to judge responsibility in:

a) A real Neuro-reductionistic world (to an abstract question)

b) A real Psychological-deterministic world (to an abstract question)

c) An alternate Neuro-reductionistic world (to an abstract question)

d) An alternate Psychological-deterministic world (to an abstract question)

e) An alternate Neuro-reductionistic world (to a concrete question)

f) An alternate Psychological-deterministic world (to a concrete question)

Let us then discuss these scenarios one by one. But before that we would describe the Neuro-reductionistic world and the Psychological-deterministic world using Nahmias' phraseology.

Neuro-reductionistic world: Imagine that the neuroscientists in our universe or in an alternate universe (which was given an imaginary name Erta) have discovered that every single decision and action we perform is *completely caused by the particular chemical reactions and neurological processes occurring in our brain* at the time, and that these chemical reactions and neurological processes in the brain are completely caused by earlier events involving our particular genetic makeup and physical environment.

Psychological-deterministic world: Imagine that psychologists in our universe or in an alternate universe (Erta) have discovered that every single decision and action we perform is *completely caused by the particular thoughts, desires, and plans we have* at the time, and that these

thoughts, desires, and plans are *completely caused* by earlier events involving their particular genetic makeup and upbringing.

We will now present the results of a) and b). 81 subjects were given the description of the (real) Neuro-reductionistic world and another 71 subjects were presented with the description of the (real) Psychological-deterministic world. They were then asked to respond to the following questions with either "Yes," "No" or "I don't know".

1) Taking the above scenario for granted, do you think we are morally responsible for whatever we do?

(2) Do you think we deserve to be given credit or blame for our actions?

Now, It was found that subjects were more inclined to perceive the real world, where choices were determined by mentalistic states like thoughts, desires etc. as conducive to holding one guilty and praiseworthy; the world where brain states were an established cause for choices was viewed far less amenable to moral accountability (see Table 9).

	The Brain World	**The Mentalistic World**
The inhabitants have MR	40.7%	88.6%
The inhabitants deserve blame	37.7%	85.7%
The inhabitants deserve praise	48.7%	85.9%

Table 9

Let us now turn to the results of c) and d). In this experiment, 90 subjects were presented with the Neuro-reductionistic world condition and 65 subjects were presented with the Psychological-deterministic world scenario. Participants on both the conditions were additionally told that these worlds are similar to our world but still differ from ours as a species called Ertans inhabit them. However, the findings by the neuroscientists (in

the Neuro-reductionistic world) and those by the psychologists (in the Psychological-deterministic world) remained the same. They were once again given the previous set of abstract questions aimed at drawing out their moral intuitions. And as can be seen in the Table below, subjects tended to be more compatibilist in the Psychological-deterministic world than in the Neuro-reductionistic world.

	The Brain World	**The Mentalistic World**
The Ertans have MR	52.4%	71.9%
The Ertans deserve blame	50.6%	70.3%
The Ertans deserve praise	67%	78.1%

Table 10

Finally, it is the turn for e) and f). Like Nichols, Nahmias also observed that a concrete description of an act censurable from a moral point of view or a morally commendable act mitigates the circumscribing effect of determinism, or as Nahmias would prefer to call, *mechanism*. In order to test the effect of concreteness of morally salient acts on judgments about their permissiveness, Nahmias presented his participants in both the Neuro-reductionistic Ertan world condition and the Psychological-deterministic Ertan world condition with an account of a morally good act (donating a large sum of money to an orphanage by an Ertan called Smith) and a morally reprehensible act (Smith killing his wife to keep alive his extra-marital relationship). Attention now would be drawn in particular to the responses to the morally condemnable act. Here Nahmias found a pattern of results that were in conformity with those in Nichols' concrete condition experiments. Subjects tended to overlook the mechanistic description of the psychological setup of the Ertans and maintained that they would be no less

culpable than if their choices were to be governed by their own intentional states. And those in the Psychological-deterministic world seemed to be ever more enthusiastic about holding the agent of the reprehensible act responsible. The responses are given in Table 11.

	Bad Act in the Brain World	**Bad Act in the Mentalistic World**	**Good Act in the Brain World**	**Good Act in the Mentalistic World**
The Ertans have MR	79.2%	81.1%	63%	68.5%
The Ertans deserve blame	74.3 %	85.6%	NA	NA
The Ertans deserve praise	NA	NA	70.5%	75%

Table 11

3. In Quest of the Origin of the Intuitional Dilemma

While combing through Nichols' work, we observed that Nichols used affect or moral sentiments (say, anger and sympathy) as a variable. And quite in accord with what he expected, affect-laden concrete conditions seemed to deflect lay intuition from taking into consideration any deterministic threat, giving rise to compatibilist responses. Incompatibilist responses however were found to be triggered by emotionally neutral scenarios. Following this, Nichols and Knobe found it plausible to posit a hybrid theory. As Roskies puts it:

> Nichols and Knobe postulate that people's conflicting intuitions in different moral scenarios are attributable to the operation of two different subsystems that govern reasoning about moral responsibility. One is harnessed in emotionally neutral cases such as the evaluation of abstract questions, which tends to produce judgments consistent with

incompatibilist intuitions, and the other is triggered by emotional responses and leads to judgments in line with compatibilist intuitions. [8]

Nichols' affect-based mechanism is somewhat reminiscent of Peter Strawson's theory of non-detached, interpersonal *reactive attitudes* which the latter claims to be insulated from any deterministic threat. According to Strawson, an array of such human emotions as anger, gratitude, forgiveness, resentment etc. that enable us to participate in a human relationship, which he has famously given the nomenclature of *reactive attitudes*, is the springboard of compatibilist intuitions. We tend to excuse ourselves from these reactive attitudes, or rather it would be better to say that we begin to review our emotion-ignited attitudes only when it comes to determining the quantum of responsibility of "only a child", or "a hopeless schizophrenic" or a "perverted" or someone who "behaved purely compulsively"- the kinds of cases that demand the employment of what he calls the *objective attitudes*. But else he makes a strong point that:

> […] it has never been claimed that as a consequence of the truth of determinism […] it would follow […] that anyone who caused an injury *either* was quite simply ignorant of causing it *or* had acceptably overriding reasons for acquiescing reluctantly in causing it.[…] [9]

Echoing Strawson's view that moral sentiments are at the heart of an affective mechanism and consequently account for compatibilist tendencies, Nichols differs from Strawson in that, he proposes the view that incompatibilist reactions are also in a way provoked by moral emotions or rather the diminishing effect of them. Thus, his view is not in tandem with Strawson's Insulationism, but with the Enshrinement Theory propounded by the likes of Galen Strawson and Derk Pereboom. The Enshrinement theorists try to show that moral sentiments also entrench and drive incompatibilist intuitions in contradistinction to Insulationism that maintains

---

[8] Roskies: 2006, p.422.
[9] Strawson: 1980, pp. 10-11.

that moral sentiments act as a bulwark against incompatibilist attitudes. In support of their claim, they draw our attention to such cases as the following:

A person called Harris engages himself in a strongly reprehensible crime like murder and is brought to book. On hearing the incident, our negative sentiments are naturally evoked. But as the trial continues, chilling stories of his turbulent past life, a bullying and uncaring family, financial suffering etc. begin to surface. The anger now starts to wear off and our reactive emotions become less pronounced. [10]

Nichols accepts the Enshrinement theory, as he finds it more tenable that incompatibilism is triggered by attenuation of moral anger.

But a difficulty seems to exist with an account of such origin of incompatibilist intuitions as proposed by the Enshrinement theorists. While the moral anger aimed at the original perpetrator diminishes and the perpetrator now becomes the cynosure of sympathy that only a victim of a violent crime can evoke, we find the new perpetrators in his family members and direct our initial moral resentment against them. But if we further find that the moral degradation of these people is also in a similar manner attributable to conditions that they had no hands on, then won't we be allured to pass on our incompatibilist feelings to yet another agent and the process would go on ad infinitum? Borrowing Dennett's words we are then urged to say that the buck has to stop somewhere.

One can nevertheless observe that Nichols successfully highlights a salient role played by affect (or the lack of it) in manipulating intuitions-compatibilist in the first case and incompatibilist in the other. On his interpretation, therefore, one possible factor responsible for all this conflict is affect; it is affect that makes all the difference. Indeed, Nichols toys with the following possibility:

---

[10] Such concrete cases as the one mentioned in order to demonstrate how even incompatibilist feelings can be harboured by emotions are developed by Enshrinement theorists like Galen Strawson and Derk Pereboom. This particular scenario can be found in Nichols: 2007b.

> Our results could be understood as a consequence of the variable involvement of emotion in the assessment of scenarios set in our own or in other worlds. One can think of alternate universes as more removed and less personally involving than our own, so that the very same scenarios would differentially involve emotional areas during processing of questions of moral responsibility. This differential involvement would explain. [11]

Nahmias, on the other hand, puts forward the claim that there is a misplaced apprehension and suspicion on the part of the folk that 1) our freedom to choose and act is overridden by physical and chemical processes in the brain- a paradigmatic case of a mechanistic phenomenon and 2) these mechanistic processes sort of reduce our power of choosing and deciding to nothing more than a *brainwork*. And it is the differentiation that they make between mentalistic processes and mechanistic processes that is responsible for all the conflicting intuitions.

To paraphrase, Nahmias is of the view that an intentional or a participant stance in line with Strawson's reactive attitudes evokes compatibilist responses whereas a mechanistic stance triggered by a fear or a *bypassing* threat that our intentional states, which we suppose to underlie our acts and decision making processes, are reduced to brain-powered, epiphenomenalistic states is responsible for incompatibilist intuitions. Nahmias introduces the notion of Mechanism Incompatibilism in contrast with Pure Incompatibilism. For him, folk as such, may not perceive any threat from determinism; what they count as antagonistic to their concept of free will and MR is a reductionistic description of themselves and their behavioural system. He, accordingly, attributes the low outcome of compatibilist responses on the Neuro-reductionistic scenario to this apprehension of reductionism. As he puts it:

> […] from philosophers to scientists to journalists to the ordinary "folk" we have surveyed—share the intuition that "if our brain makes us do

[11] Nichols (forthcoming), p.9.

> it, then we aren't morally responsible". We think that this intuition runs deep and that it is driven by people's tendency to view a reductive, mechanistic explanation of behavior—for instance, in the neuroscientific language of neural processes and chemical reactions—as inconsistent with a mentalistic (or intentional) explanation—in the psychological language of thoughts, desires, and plans. Because people also tend to ascribe free will (FW) and moral responsibility (MR) only to agents whose actions can be understood in terms of their mental states, people tend to see reductive mechanism as incompatible with FW and MR.[12]

Further, Nahmias seems to advance this view with all the more enthusiasm as in the Fred and Barney case as well as in the Supercomputer scenario participants were found to give a very lukewarm response to genetic determinism and physical law determinism respectively. However, having said that, he adds:

> […] that the claim that incompatibilism is intuitive to ordinary people rests on a failure to distinguish 'pure' incompatibilism (between determinism *per se* and free will) and 'derivative' incompatibilism (between deterministic *reductionism* and free will).[13]

But one might ask whether mentalistic notions are perpetually at loggerheads with mechanistic notions. Dennett once said that:

> The Intentional stance toward human beings, which is a precondition of any ascription of responsibility, *may* coexist with mechanistic explanations of their motions.[14]

We will, however, not enter into the arguments that Dennett subsequently offered, as it does not come within the purview of this paper. But, we can note that concurring with Dennett, Nahmias also says that

---

[12] Nahmias: 2007, pp. 215-216.
[13] Nahmias: 2006, p. 230
[14] Dennett: 1982, p. 170

mechanistic system can also be purposive and, intentional systems. But, he has reason to believe that unprimed intuitions are not directed by such a belief, may be because our experience does not warrant that. Thus he says:

> […] when people adopt the mechanistic stance toward an agent (for instance, when primed by a description of decision-making in terms of neural processes), then they tend to disengage from the participant stance. And they tend to treat the mechanistic explanations as precluding mentalistic explanations. [15]

It may seem at this juncture, that Nichols and Nahmias are explaining the anomaly in intuitions from two very different expository frameworks. But signs of reconciliation, nevertheless, can be traced in both the positions. Nahmias, for instance, acknowledges the role of emotions in galvanizing judgments of MR, especially in accounting for those cases where despite a portrayal of a reductionistic description of human acts, compatibilist responses do not exactly put up a poor show (see Table 10 and Tale 11). He, however, seems to be more a supporter of an affective competence model. For him, emotional responses should be considered *enabling factors* that engage the cognitive processes that we employ from within the participant or intentional stance. Although he grants the possibility that the competence of affect may suffer a setback; that these emotion-driven cognitive processes may function in a sub-optimal way when we make abstract judgments about agents in general conditions.[16]

Again, like Nahmias, Nichols also suspects the "natural-ness" of incompatibilist responses, at least the kind found by Nahmias. He avers:

> The idea that our behavior is not caused by our mental states is truly, deeply disturbing. […] if our actions aren't caused by our mental states, then commonsense psychology is profoundly mistaken. We think that our actions are caused by what we intend, and our intentions are

[15] Nahmias: 2007, pp.233-234.
[16] Nahmias: op. cit., p.235.

produced by our thoughts and wants. Epiphenomenalism trashes all of this.[17]

4. Postscript

The presentation of the folk-study on moral reasoning by Nichols and Nahmias, the two aficionados and champions of XP (experimental philosophy as it is fondly called) in tow, in the foregoing, has the following objective:

Laying bare the areas where philosophers irrespective of whether they cling on to the compatibilist view or to the Libertarian (agent-causationist as well as non-agent causationist) standpoint can go wrong and thus alerting them to the exercise of "exercising some temperance", so to speak, even as they claim that their view about free will and MR is more intuitive. In fact Nichols and Nahmias both form a consortium of sorts in sharing the view that this descriptive project of plumbing folk intuitions and drawing a parallel between folk beliefs concerning choices and responsibilities and the rationales concerning the same, ambitiously put forward by their philosophical counterparts has an enormous bearing on the normative or prescriptive question. They certainly believe *a fortiriori* that work on meta-ethical issues and practical moral philosophy will be enlightened, given the wealth of data on the asymmetric nature of folk predispositions about the issues of moral accountability they have garnered. Now, if the folk display a wavering attitude, when it comes to ascertaining culpability, in the light of a circumscribing portrayal of our biological and psychological makeup, then do we need to rethink and revamp our present moral practices of reward and retributive punishment? Both, Nichols and Nahmias point out the importance of the findings of their empirical research in addressing this normative or prescriptive question. The normative question also gives rise to two warring camps - that of the Revisionists or Revolutionists versus the Conservatists. Revolutionists maintain that we need to embark on a thorough review of the existing moral practices lest the

[17] Nichols: 2006, p.310.

folk intuitions about MR turn out to be a distorted case of moral judgment. The supporters of Conservatism, on the other hand, believe in holding on to the moral practices. Focusing on the debate, however, should better be left for another occasion. We may, nonetheless observe that whether it is the affective competence or affective bias or a natural participant attitude as fomented by the reactive attitudes driving compatibilist responses, assigning responsibility is a task that involves our emotionally intertwined practical experience. Hence, perhaps there is no immediate need for any Revisionism.

References

Dennett, Daniel. (1982). "Mechanism and Responsibility" in Gary Watson (Eds.), *Free Will*. New York: Oxford University Press.

Dennett, Daniel. (1984). "I Could Not Have Done Otherwise-So What?" in *The Journal of Philosophy*, 81(10), pp. 553-565.

Ekstrom, Laura. (2002). "Libertarianism and Frankfurt-Style Cases" in Robert Kane (Eds.), *The Oxford Handbook of Free Will*. New York: Oxford University Press.

Nahmias, Eddy, Morris, Stephen, Nadelhoffer, Thomas, Turner, Jason. (2005). "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility" in *Philosophical Psychology*, 18(5), pp. 561–584.

Nahmias, Eddy. (2006). "Folk Fears about Freedom and Responsibility: Determinism vs. Reductionism" in *Journal of Cognition and Culture*, 6(1-2), pp. 215-237.

Nahmias, Eddy, Coates, D. Justin, Kvaran, Trevor. (2007). "Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions" in *Midwest Studies in Philosophy*, XXXI, pp. 214-242.

Nichols, Shaun, Roskies, Adina. "Bringing Moral Responsibility Down to Earth**"** (forthcoming) in *Journal of Philosophy*.

Nichols, Shaun, Knobe, Joshua. (2007a). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions" in *Nous*, 41(4), pp. 663-685.

Nichols, Shaun. (2007b). "After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes" in *Philosophical Perspectives*, 21(1), pp. 405-428.

Roskies, Adina. (2006). “Neuroscientific Challenges to Free will and Responsibility” in *Trends in Cognitive Sciences*, 10(9), pp. 419-423.

Strawson, Galen. (1986). *Freedom and Belief*. Oxford: Oxford University Press.

Strawson, Peter. (1980). “Freedom and Resentment” in *Freedom and Belief*. Oxford: Oxford University Press.

Wolf, Susan. (1990). *Freedom within Reason*. New York: Oxford University Press

Does Moral Discourse Require Robust Truth?

Fritz J. McDONALD [*]
Oakland University

Abstract:

It has been argued by several philosophers that a deflationary conception of truth, unlike more robust conceptions of truth, cannot properly account for the nature of moral discourse. This is due to what I will call the "quick route problem": There is a quick route from any deflationary theory of truth and certain obvious features of moral practice to the attribution of truth to moral utterances. The standard responses to the quick route problem are either to urge accepting a conception of truth more robust than deflationism (Boghossian 1990), or to revise deflationary accounts in order to block straightforward attribution of truth to moral utterances (Field 1994). I contend that neither of these standard responses is well-motivated, for it is a merit of deflationary accounts rather than a defect that such accounts present a quick route to moral truth.

Keywords: Deflationism, Truth, Realism, Antirealism, Metaethics, Expressivism

It has been argued by several philosophers that a deflationary conception of truth, unlike more robust conceptions of truth, cannot properly account for the nature of moral discourse. This is due to what I will call the "quick route problem": There is a quick route from any deflationary theory of truth and certain obvious features of moral practice to the attribution of truth to moral utterances. Due to this quick route from deflationism to moral truth, any such deflationary conception is supposed to

[*] E-mail: fritzmc@gmail.com.

preclude the possibility of formulating an "anti-realist," "expressivist," "projectivist," or "non-factualist" account of ethics. Whether any one of these meta-ethical "isms" is a correct or incorrect view should, these philosophers contend, be a matter for serious debate, not a matter settled in a trivial manner by a theory of truth.

Focus on the question of whether or not deflationary accounts of truth rule out certain meta-ethical theories has resulted in insufficient attention being given to the question of how coherent the account of moral truth purportedly essential to these "isms" actually is, and whether this account of moral truth could be accepted without doing a great deal of harm to our commonsense approach to moral talk as well as our philosophical theories of morality.

In this paper, I will briefly review the features and advantages of three prominent deflationary theories of truth: disquotationalism, prosententialism, and the minimalist theory. Then I will show how each of these theories is open to the quick route problem. Much of the attention in the recent literature has been focused on the minimalist theory, but it ought to be noted that the quick route problem is not only the minimalist's burden. The standard responses to the quick route problem are either to urge accepting a conception of truth more robust than deflationism (Boghossian 1990), or to revise deflationary accounts in order to block straightforward attribution of truth to moral utterances (Field 1994). I contend that neither of these standard responses is well-motivated, for it is a merit of deflationary accounts rather than a defect that such accounts present a quick route to moral truth.

Inflationism and Correspondence

It is worth noting that there are quite good reasons to accept a deflationary conception of truth that stand apart from the considerations related to moral discourse raised in this paper. An inflationary theory is a theory that identifies truth with a property, and then provides an analysis of the underlying nature of that property. On the paradigmatic inflationary account, the correspondence theory of truth, the property of being true is identified with the property of being a proposition that corresponds to the

facts. This rough characterization obviously requires further characterization itself. Spelling out what a fact is, and what it would be for a proposition to correspond to such a thing, is a major task of the correspondence theory. Exactly what a fact is and what it would mean for a proposition to correspond to a fact is obscure. Is a fact some sort of sentence-shaped object in the world? Is the correspondence some kind of resemblance? What are the criteria of identity for facts? Can a correspondence theorist avoid the notorious slingshot argument (Davidson 1984)? Is it the case that if a sentence corresponds to any fact at all, it corresponds to all facts?

In order to clarify some of these obscurities, philosophers influenced by Hartry Field's classic paper "Tarski's Theory of Truth" have attempted to spell out the nature of this correspondence not in terms of facts but in terms of the reference of subsentential units. These referential relations are then, in turn, explained in terms of a causal theory of reference. The philosophers carrying out this project have run into serious difficulties. How does one assure that the causal links between a term such as 'Earth' and the Earth itself are specified in the proper way to explain reference? How does a causal theorist of reference distinguish appropriate causal chains from inappropriate causal chains? Complex theories presented by philosophers such as Fred Dretske (1988) and Jerry Fodor (1990) attempting to specify the proper causal link between words and the world have been found lacking[1].

The case of moral truth raises a particular worry for the correspondence theorist, especially the correspondence theorist who appeals to a causal theory of reference. If such a philosopher were inclined to think moral utterances are capable of being true or false, that philosopher would be forced into accepting what would be widely regarded to be an implausible account of the metaphysics of morality. Only a certain group of philosophers, the naturalistic "moral realists," think that moral properties are properties that figure into causal relations (Boyd 1988, Railton 1986).

---

[1] Barry Loewer (1987) details the problems presented by misrepresentation for Dretske's theory, and Fred Adams and Ken Aizawa (1994) raise serious problems for Fodor's attempted resolution of the misrepresentation problem.

Philosophers who hold "expressivist" and "constructivist" views, as well as many others impressed by arguments dating back to Hume (2000) and Moore (1903), are skeptical of the claim that moral properties such as goodness, virtue, and justice are invoked in causal explanations and laws. Any correspondence theorist inclined to accept that moral utterances can be true or false would have to answer the Humean, Moorean arguments against naturalism, a major burden. Thus it is worth noting here that it is the inflationary, correspondence account that would be in conflict with a range of meta-ethical views that follow Hume and Moore in rejecting naturalistic moral realism.

Deflationary Theories

Philosophers who have been skeptical of the analyses given by inflationary theories such as the correspondence theory have asked whether it is a mistake to assume that there is a property of truth with a substantial underlying nature. Is it necessary to give such an account of the property of truth in order to explain the function of the predicate 'true'? Or is there a different account that fully explains the function of this predicate? If one can give a full account of the function of the truth term without reference to one of these vexed, incomplete theories of the underlying nature of truth, why would any further theorizing be required?

The summaries I will present below will, I hope, make it clear that regardless of the application of the label 'deflationist' to these theories, there are significant differences among these theories of truth. Each of these theories takes a different position on the role of the truth predicate. These theories differ on the issue of whether or not truth is a property. Some of these theories involve complexities such as appeals to substitutional quantification, whereas others do not. It is not uncommon to find objections raised against particular deflationary theories that involve ignorance of the difference between one deflationary theory and another. For instance, the minimalist theory is sometimes criticized for denying that there is a property of truth, even though Horwich (1998b) quite clearly claims that minimalism holds that there is a property of truth.

While these theories may differ in important respects, there are certain features these theories have in common that give rise to the issues I will discuss in this paper. As I will explain in detail in the sections following the summaries, all of these theories share the feature of trivializing the distinction between asserting that p and asserting that p is true. This shared feature, as noted above, plays a significant role in the debates over truth and its relation to meta-ethical controversies.

The Disquotational Theory of Truth

Unlike a Tarskian account of truth, the disquotational theory of truth proposed by W.V. Quine (1970, 1992) and Field (1986) does not require that truth be accounted for in terms of satisfaction, denotation, and recursive rules for sentence construction. Rather, this account claims, as Quine puts it, that "truth is disquotation" (Quine 1992, 80). For any sentence in a language, 'p' is true iff p. One can appeal to truth in order to disquote the sentence mentioned on the left-hand side of this biconditional. This fact, also noted by the Tarskian theory, is regarded as basic on the disquotational account, requiring no further explanation. Anyone with a grasp of the notion of truth will understand that D1 and similar instances of the disquotational schema are acceptable:

D1: 'The Earth moves' is true iff the Earth moves.

The disquotational theory, unlike the minimalist account (discussed below), does not appeal to propositions as the vehicle of truth. The vehicle of truth on the disquotational account is a class of sentences, the eternal sentences. Eternal sentences are context-independent sentences. This requirement is a significant one for this account. For instance, it would be troublesome if the left side of the following schema instance were read relative to one context (say, July 21, 2004) and the right side of the biconditional were read relative to another (July 20, 2004):

D2: 'It is Tuesday today' is true iff it is Tuesday today.

Thus the only proper candidate sentences for instances of the disquotational schema are eternal sentences, sentences with truth values not dependent upon context, such as 'July 20, 2004 is a Tuesday.' The

attribution of truth to sentences only, and not to propositions, is appealing to philosophers who are dubious of the existence of propositions.

Unlike a redundancy theory of truth, the disquotational account does not assume that the notion of truth plays no significant role in the language. On this account, it is correctly noted that truth plays the important role of allowing one to formulate generalizations about true sentences. For instance, it would be impossible for a redundancy theorist to account for the fact that all sentences with the logical form p∨~p are true. One can only assert particular instances of this schema, p∨~p, and attribution of truth to particular instances of this schema are eliminable redundancies. The disquotational account allows one to semantically ascend from each instance of the schema p∨~p to the metalinguistic level. Take particular instances such as: 'the Yankees will win the World Series ∨ the Yankees will not win the World Series' and 'the Red Sox will win the World Series ∨ the Red Sox will not win the World Series'. We can then ascend, via the disquotational schema, to the metalinguistic level to assert that 'The Yankees will win the World Series ∨ the Yankees will lose the World Series' is true, along with all of the other instances of this schema. Such semantic ascent allows one to assert that the conjunction of all instances of this schema p∨~p are true. By allowing for the construction of such infinite conjunctions, the disquotational account explains the important role played by attributions of truth.

The disquotational account only explains one class of attributions of truth, namely attribution of truth to sentences. How are we to explain other attributions of truth, such as attribution of truth to beliefs? One possible way to do so is to claim that truth is attributed to the propositions expressed by these beliefs. However, if a disquotational theorist appeals to propositions, then there is no significant difference between this account and the minimalist theory discussed below.

A merit of minimalism that is not shared by the disquotational theory is that the disquotational theory cannot explain how we can apply the notion of truth to sentences that we do not understand. A speaker can only comprehend instances of the schema spelled out in her own language; A

monolingual English speaker would not know why it is that 'Schnee ist weiss' is true iff *schnee ist weiss*. For this reason, Field restricts the theory of truth to a specific set of utterances, "only…utterances a person understands" (Field 1994, 405). This limitation in the ability of the disquotational theory to explain the concept of truth—limiting the concept to one that only applies to the utterances one understands—is a consequence of the disquotational theorist's refusal to countenance propositions. Without such restrictions, other theories such as minimalism can avoid this limitation of the disquotational theory.

The Prosentential Theory of Truth

The prosentential theorist, like the disquotational theorist, provides an account of the role of the truth predicate in a language. The most significant difference between the prosentential theory and all of the other deflationary theories of truth is the distinctive account the prosentential theorist gives of the role played by the truth predicate. The prosentential theory claims that assertions such as 'That is true' have a function analogous to pronouns.

On one reading of sentence AP, the pronoun 'he' is an anaphoric pronoun:

AP: Derek knew that he needed to hit a home run.

The pronoun 'he' has the same referent as its antecedent, the name 'Derek.' It obtains this referent by being anaphorically dependent upon the antecedent. In addition to anaphoric pronouns, as Grover, Camp, and Belnap (1975) point out, there are anaphoric proadjectives, such as 'so' in the following quotation from Alexander Pope: "To make men happy and to keep them so" (Grover, Camp, and Belnap 1975, 84). The expression 'so' essentially plays the same role in this sentence as a second occurrence of 'happy' would play, describing how the men being discussed by Pope are kept. The word 'so' inherits its meaning from its antecedent, 'happy.' In the following discourse, DIS1, 'That is true' is, according to the prosentential theory, a prosentence:

DIS1:

Galileo: The Earth moves.

Castelli: That is true.

The prosentence 'That is true,' asserted by Castelli, is anaphorically dependent upon its antecedent, Galileo's assertion 'The Earth moves.' Just as the pronoun 'he' in AP inherits its content from its antecedent and the proadjective 'so' in the Pope quotation inherits its content from the adjective 'happy,' according to prosententialism the prosentence 'That is true' has the same content as its antecedent. Thus, in this context, 'That is true' means that the Earth moves.

The prosentential theory has to contend with one of the difficulties that plagued the redundancy theory of truth. The bare-bones prosentential theory summarized above does not have the resources to explain the meaning of sentences such as K1:

K1: What Kerry said about the Iraq war is true.

In order to explain such occurrences of the truth term, Grover, Camp, and Belnap (1975) claim that the English sentence K1 is equivalent to the sentence K2:

K2: For each proposition regarding the Iraq war if Kerry said that it is true then it is true.

As Paul Horwich has pointed out, in order to explain why K1 and K2 are equivalent, one would appeal to the fact that 'true,' *pace* the prosentential theory, is a genuine logical predicate. For K1 is equivalent to saying:

If Kerry said, regarding Iraq, that p, then p.

We can then use the minimalist theory (detailed below) to expand this into:

If Kerry said, regarding Iraq, that <p> is true, then <p> is true.

This expansion is equivalent to K2, but the explanation of how we derived K2 from K1 relies on the resources of the minimalist theory, and, as noted above, relies on considering 'true' a genuine predicate.

The Minimalist Theory of Truth

Minimalism is the view that the meaning of the term 'true' in English (and the meanings of similar terms in other languages) is best

analyzed in terms of a fact regarding the use of the term by speakers of the English language. The meaning of 'true' is explained fundamentally by the acceptance of a trivial schema T:

T: <p> is true iff p.

In the schema, '<p>' is short for 'the proposition that p.'

Speakers of English are inclined to accept, for any given proposition, <p>, that the proposition that p is true iff p. According to minimalism, the fact that speakers accept instances of such a schema explains the purpose of the notion of truth, which is to allow one to form generalizations such as 'Everything the president said in his speech was true' and 'All instances of 'if p, then p' are true.' The generalizing role of truth is the sole purpose of the notion of truth. No further facts, beyond acceptance of the schema, are required in order to specify the meaning of the term 'true.'

For reasons that I have discussed above in the sections on the competing deflationary theories of truth, minimalism has a number of advantages over its competitors. Regardless of these differences, as I have noted, there is one key similarity between all deflationary theories. Each deflationary theory makes trivial the distinction between asserting that p and asserting that it is true that p. Whether or not this trivialization of this distinction is troubling will be considered below.

Attribution, Denial, Anomaly

An extensively discussed question in the philosophical field of meta-ethics, primarily in the 20th and early 21st century, is whether utterances pertaining to normative matters generally and moral matters specifically are capable of being either straightforwardly true or straightforwardly false. This question has been raised due to the view, held by many philosophers, that there a significant difference between the class of nonnormative and nonmoral utterances, such as 'The Earth moves' and 'Albany is the capital of New York State' and the class of normative and moral utterances such as 'Rape is wrong,' 'Great inequalities in the distribution of wealth are unjust,' and 'One should be polite in the company of strangers.' This view is due to the wide-spread belief that a characterization of the semantic difference between moral/normative discourse and nonmoral/nonnormative discourse

is required in order to characterize the difference between moral/normative matters and nonmoral/nonnormative matters.

One could take one of several positions in response to aforementioned question regarding the truth or falsehood of moral and normative utterances. One could hold that such utterances are either true or false, and therefore in this respect do not differ from the nonmoral and nonnormative utterances. One could hold that such utterances are neither true nor false, and thus differ from the nonmoral and nonnormative utterances in this respect. Or, one could hold that such sentences are capable of being true or false, but are true or false in some distinctive way that indicates the difference between normative/moral utterances and nonnormative/ nonmoral utterances.

One way to spell out this third option, to offer a distinctive kind of truth attribution to these utterances, would be to hold a relativist view. Such a view holds that normative and moral utterances are true or false only relative to a particular individual or social perspective. Another way to spell out such a view would be to claim that the theory of truth for normative and moral utterances differs from the theory one would give for other utterances, as Wright (1992) claims.

In order to simplify the subsequent discussion of this issue and avoid repetition, I will use the following terms to refer to the theses discussed in the previous paragraph. I will call the approach that allows for the straightforward attribution of truth and falsehood to normative and moral utterances the Attribution Thesis. The view denying that truth and falsehood can be attributed to normative and moral utterances will henceforth be called the Denial Thesis. Finally, the theories calling for truth and falsehood of a distinct kind to be attributed to normative and moral utterances will be called instances of the Anomaly Thesis.

The Quick Route to Attribution

On a deflationary theory of truth, including any of the three deflationary theories discussed above, it would seem that there is a fairly quick route from the fact that people make sincere moral assertions to the Attribution Thesis. Using the resources of any of these theories, one can

show that the inference from an assertion that p to the assertion that p is true is a trivial one. Thus as soon as one commits oneself to holding that rape is wrong, one would also commit oneself to hold (if one has a grasp of the notion of truth) that it is true that rape is wrong.

I will show that there is such a quick route on any deflationary theory of truth, and I will offer several arguments for regarding this as a merit of these deflationary theories rather than a defect.

The Quick Route: The Disquotational Theory

The disquotational theory, as I mentioned above, accounts for truth without making appeal, as a Tarskian theories does, to principles regarding predicate satisfaction, denotation, and the role of connectives. It does, however, regard T schema instances along the lines of TR as basic:

TR: 'Rape is wrong' is true if and only if rape is wrong.

Thus, given that 'Rape is a wrong' is a meaningful sentence, the disquotational schema can be applied to a sincere assertion of 'Rape is wrong' to show that truth ought to be attributed by that person to the sentence 'Rape is wrong.'

The Quick Route: The Prosentential Theory

For the prosentential theorist, the quick route from sincere moral assertion to the Attribution Thesis is illustrated by the fact that purported prosentences can be and often are used in contexts where moral assertions are the antecedents of such prosentences. Thus if Larry asserts that 'Great inequalities in the distribution of wealth are unjust,' and Barry responds 'That is true,' what does Barry's assertion mean? On the prosentential account, as noted above, Barry's utterance is anaphorically dependent upon Larry's utterance, and thus his assertion 'That is true' has the same meaning as 'Great inequalities in the distribution of wealth are unjust.' There is no reason to think that we cannot use the resources of the prosentential theory to form prosentences anaphorically dependent upon moral and normative utterances in just the same way we use these resources to form prosentences anaphorically dependent upon nonmoral and nonnormative utterances.

The Quick Route: The Minimalist Theory

On the minimalist theory, unlike some theories such as disquotationalism and prosententialism, truth is attributed not directly to sentences, but rather to propositions. So, in order to establish that there is a quick route from sincere assertions to the Attribution Thesis on the minimalist theory, we first have to ask whether we should regard moral utterances as assertions that involve the expression of propositions.

On the assumption that an utterance such as 'Rape is wrong' expresses a mental state with propositional content, we would take the utterance to express the proposition that rape is wrong. Is there a reason to reject the claim that the mental state expressed in this situation does express such a proposition? Are such utterances not meaningful? Do we not use them in all of the ordinary contexts in which we also use meaningful utterances? The *prima facie* correct view is that moral utterances do in fact express propositions.

There is some historical precedent for rejecting the *prima facie* view that moral utterances express propositions. A.J. Ayer, in his account of moral language in *Language, Truth, and Logic*, claims that moral utterances do not express propositions for they are unverifiable. Given that there are no propositions expressed by such utterances, and truth is a property of propositions, then moral utterances are not capable of being true or false[2]. The contemporary deflationist need not accept the verificationist commitments of Ayer's account, hence the deflationist ought not to claim that moral utterances fail to express propositions.

If an utterance of 'Rape is wrong' does express the proposition that rape is wrong, then the following would be an instance of the minimalist truth schema:

TR: <Rape is wrong> is true iff rape is wrong.

---

[2] Boghossian claims that Ayer fails to notice a tension between an emotivist account of ethics and a redundancy theory of truth. This argument of Boghossian's—a quite influential argument—overlooks Ayer's verificationist account of meaning and the role it plays in blocking attribution of truth to moral utterances.

Thus, any speaker with an understanding of the notion of truth, according to the minimalist theory, would be able to recognize that the claim that it is true that rape is wrong is a consequence of the truth schema and that rape is wrong. Thus, on the minimalist theory, there is a quick route from asserting that p to asserting that p is true.

Should We Avoid the Quick Route?

If these deflationary theories, along with our practice of moral discussion and argument, give us good reason to affirm the Attribution Thesis, should we regard this as a bad thing? Without delving into the complex details of specific meta-ethical theories, there are several arguments that can be put forward for regarding this quick route to the Attribution Thesis as the appropriate road to take.

First, normative and moral assertions have all of the same surface features as nonnormative and nonmoral assertions. 'Killing is wrong' appears to attribute wrongness to killing in just the same way that 'The Earth is round' attributes roundness to the Earth. Taking this surface structure into consideration provides a *prima facie* reason for regarding this moral utterance as similar in other respects to nonmoral utterances.

As I will discuss below, some philosophical theories offer reasons to believe that this surface appearance is misleading. These theories claim that apparent moral predications are not genuine predications. On these theories, moral predicates are like "sakes," a well known example discussed in Quine 1960. If I were to say that "I am doing this for Susan's sake," it would be a mistake to think that this sentence involves reference to some strange kind of entity, a sake. An analysis of the meaning of this sentence will show that no reference to sakes is required—what the sentence really means is that I am doing this in order to help Susan.

A question that must be asked about any theory of the meaning of moral terms that attempts to explain away the surface appearance that predicates such as 'right,' 'wrong,' and 'just' are genuine predicates is whether it is reasonable to attribute such a theory to ordinary speakers of a language. One would suspect that no speaker of English takes seriously the apparent reference to sakes in the examples discussed above, and this fact is

reflected in our use of the term. Is this really the case with the moral predicates? Is a semantics for moral predicates that regards them as something other than genuine predicates really implicit in ordinary practice?

It is quite clear that moral predicates, unlike 'sake,' do not fit Quine's description of a defective noun. Defective nouns, nouns that function as 'sake' does, have the following features according to Quine:

> …we never use 'sake' as antecedent of 'it,' nor do we predicate 'sake' of anything. 'Sake' figures in effect as an invariable fragment of a proposition 'for the sake of,' or 'for 's sake' (Quine 1960, 236).

Moral terms and predicates do not fit this description. They do not appear only in a restricted sort of context, or within certain idiomatic constructions. Nouns that purportedly denote moral properties can serve as the antecedent to pronouns:

AN: Lester cares a great deal about economic justice, but his brother Chester could care less about it.

And, as has been discussed in this section, it is quite common to find 'good' or 'just' or other normative and moral terms predicated of things and acts.

The opponent of the Attribution Thesis may attempt to defend her view by holding that the ordinary speaker of the language is mistaken to regard apparent moral predicates as genuine. One could do this by offering an account that is not a descriptive account of our ordinary practice, but rather a revisionary one, one that tells us what sort of linguistic practice we ought to have. It is important to note here that any motivation to take such a revisionary route would, for reasons given above, have to be derived from a variety of considerations not related to our actual linguistic practice, such as metaphysical and psychological qualms.

In addition to sharing surface features with nonmoral discourse, attribution of truth to moral utterances is required to account for the role such utterances play in arguments. Take an argument such as:

P1: If murdering innocent people is always wrong, then murdering a small group of innocent people to save the lives of a larger number of innocent people is wrong.

P2: Murdering innocent people is always wrong.

C: Murdering a small group of innocent people to save the lives of a larger number of innocent people is wrong.

Such an argument certainly appears to have the form of a valid argument, an instance of *modus ponens*. However, if we reject the Attribution Thesis and hold the Denial Thesis, one could not appeal to the truth of these claims and the form of the argument to explain why the truth of these premises would lead, necessarily, to the truth of the conclusion. Acceptance of the Denial Thesis would be tantamount to claiming that there is no possibility of valid moral argument.

If some form of the Anomaly Thesis is held, there will also be serious difficulties in accounting for the validity of moral arguments. For, if the moral statements contained in moral arguments are true in some different way from the nonmoral statements contained in nonmoral arguments, then we need to appeal to a distinct notion of validity that will reflect this distinct kind of truth. Perhaps arguments containing moral statements are valid in some different way from arguments that contain only nonmoral statements. This raises further perplexities. Are arguments containing both moral and nonmoral statements valid in one way, the other, or both? A merit of the Attribution Thesis is that, by attributing truth to normative and moral utterances of the same variety as the truth that is attributed to nonnormative and nonmoral utterances, this thesis requires no revision of our ordinary notion of validity.

That this is so would be a very important result for the philosophers who have offered substantive solutions to the Frege-Geach embedding problem. The very basis for this problem is the worry that one cannot account for the validity of arguments containing moral statements. As G.F. Schueler (1988) points out in his criticism of Blackburn's response to the embedding Problem, validity is a matter of the truth of the premises necessitating the truth of the conclusion. What purpose would account for validity be without the attribution of truth to moral statements? What would

the purpose of the accounts offered by expressivists such as Blackburn (1984, 1988) and Gibbard (1990) be if moral arguments were not in fact genuinely valid ones? Isn't a solution to the Frege-Geach problem intended to show why a moral argument is in fact valid and not an instance of the fallacy of equivocation?

An analogous difficulty for the Denial Thesis is that in order to have a notion of moral knowledge that accords well with our ordinary practice, we need to attribute truth to moral utterances. As Matthew Chrisman notes in his review of Gibbard's *Thinking How to Live*, 'know' is a factive verb—One cannot claim to know a proposition unless that proposition is true. The Denial Thesis

> says that normative sentences are neither true nor false, so if that's true, we cannot have normative knowledge. Yet in a Moorean vein, one might reasonably think: "Whatever I may or may not know about the semantics of normative language, I damn well know that torturing children is wrong." (Chrisman 2005, 408).

Thus the Denial Thesis creates a tension with our commonsense conception of moral knowledge, whereas the Attribution Thesis does not.

Two further reasons for deflationists to assert the Attribution Thesis concern the nature of the deflationary theories themselves, and some theoretical considerations regarding the formulation of deflationism. The deflationist is attempting to give a theory that will capture the ordinary speaker's notion of truth. Also, the deflationary theories are simple and elegant as originally stated. I will discuss each of these considerations in turn.

One of the central aims of deflationary theories is to avoid the difficulties that have plagued previous attempts to define truth by offering an account that is based not on an analysis of the underlying nature of the property of truth, but rather on the use of the truth term by ordinary speakers of a language. These characterizations of the use of the truth term are clearly different on the various deflationary theories, but each theory essentially has the same goal: to give a correct description of the use of the

truth term. In giving such a description, these theories are not attempting (as noted above) to give a revisionary account of the practice of attribution of truth to sentences or propositions. The deflationists are not trying to explain a notion of truth that is only grasped by philosophers after consideration of a wide range of metaphysical and epistemological issues.

The philosophers who have advocated the Denial and Anomaly Theses, on the other hand, have different goals and different approaches to characterizing the notion of truth. The central reasons for holding Denial and Anomaly Theses regarding truth in some area of discourse are typically complex philosophical ones. It would be implausible, of course, to attribute an implicit grasp of such philosophical doctrines to the ordinary speaker of the language. Yet, if the proponents of the Denial and Anomaly Theses are correct, the only proper notion of truth is a notion of the sort that can be grasped only after consideration of such doctrines.

Also, as I stated above, among the theoretical merits of the deflationary theories is (to a varying degree among the theories) simplicity and elegance. If the deflationist is moved by the kinds of arguments that have been used to motivate the Denial and Anomaly Theses, then a revision in the deflationist theory will be required. The deflationist will be forced either to concede that the deflationary theory is only partially correct, and another theory of the truth of normative and moral discourse is required to fully characterize truth. Or perhaps the deflationist will add qualifications to the original theory in order to either block attribution of truth to moral and normative utterances or indicate that truth of a different sort is being attributed to these utterances. To do so would require reducing the simplicity of such theories, hence removing one of the appeals of such theory. The deflationist would be better off affirming the Attribution Thesis. In light of the arguments presented above, there are a number of consideration from outside the deflationary theories themselves that would give the deflationist good reason to accept the Attribution Thesis.

The Purported Conflict with Expressivism

Perhaps it may still be thought that there is a conflict between deflationist theories and prominent meta-ethical theories such as

expressivism[3]. A philosopher who would insist this would miss the fact that the central commitments of expressivists, historically, have been to points quite distinct from the matter of truth. Expressivists have been motivated by a metaphysical view, the view that there are no robust moral facts and properties in the world, no properties of goodness, rightness, and justice that would figure into our fundamental causal-explanatory story. They are also motivated by a view of motivation, according to which a desire or desire-like mental event is required in order to explain motivation. In light of this account of motivation, they present an account of meaning according to which the meaning of moral terms is explained in terms of the expression of a mental event such as a desire, an emotion, or what Gibbard calls "expressing a norm." None of these points has anything essentially to do with truth, and any deflationary conception will not require a philosopher to accept or reject any of these essential commitments of expressivism. It is important to reiterate, however, that there would be a tension between the expressivist account of moral facts and properties and a robust conception of truth. For on an expressivist account, there would be no robust facts to which moral utterances would correspond, hence the expressivist correspondence theorist would be forced to reject a commonsense account of moral talk and practice.

That there is no conflict between expressivism and deflationism has been recognized in a recent book by the prominent expressivist philosopher Allan Gibbard. Gibbard suggests that the option is open to the expressivist to accept minimalism, and attribute truth to moral utterances:

> Suppose instead that minimalists are right for truth, and for facts, and for beliefs: there is no more to claiming "It's true that pain is bad" than to claim that pain is bad; the fact that pain is bad just consists in pain's being bad; to believe that pain is bad is just to accept that it is. Then it's true that pain is bad and it's a fact that pain is bad—so long as, indeed, pain is bad. I genuinely believe that pain is bad, and my expressivistic theory,

---

[3] The arguments to the effect that there is such a conflict are in Boghossian 1990 and Wright 1992. These arguments have been criticized by Horwich 1993 and Hawthorne and Price 1996.

filled out, explains what believing this consists in (Gibbard 2003, 182-183).

If the arguments in this paper for attributing truth to moral utterances are correct, then expressivists ought not only consider the possibility that a deflationist account of truth is correct: it would be compulsory for such philosophers to reject any robust account of truth that would require them to reject common sense regarding moral practice and argument. Rather than being a problem for deflationism, as is widely thought, it is one of the many benefits of a deflationist account that it allows us to present such an account of morality.

References

Adams, Fred and Ken Aizawa. 1994. "Fodorian Semantics." In Stich, Stephen P. and Ted A. Warfield 1994.

Ayer, A.J. 1936. *Language, Truth, and Logic*. London: Dover.

Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Clarendon Press.

___. 1988. "Attitudes and Contents." *Ethics* 98: 501-17.

Boghossian, Paul. 1990. "The Status of Content." *Philosophical Review* 99: 157-84.

Boyd, Richard. 1988. "How to be a Moral Realist" in Darwall, Gibbard, and Railton 1997.

Chrisman, Matthew. "*Thinking How to Live*." *Ethics* 115: 406-411.

Darwall, Stephen, Allan Gibbard, and Peter Railton. 1997. *Moral Discourse and Practice: Some Philosophical Approaches*. New York: Oxford University Press.

Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Dretske, Fred. 1988. *Explaining Behavior*. Cambridge, Mass.: MIT Press.

Field, Hartry. 1972. "Tarski's Theory of Truth." In Field 2001: 3-29.

___. 1986. "Correspondence Truth, Disquotational Truth, and Deflationism." In Lynch 2001: 483-503.

___. 1994. "Disquotational Truth and Factually Defective Discourse." *PhilosophicalReview* 103: 405-52.

___. 2001. *Truth and the Absence of Fact*. Oxford: Clarendon.

Fodor, Jerry. 1990. *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Oxford: Oxford University Press.

___. 2003. *Thinking How to Live.* Cambridge, Mass.: Harvard University Press.

Grover, Dorothy. 1992. *A Prosentential Theory of Truth*. Princeton: Princeton University Press.

Grover, Dorothy, Joseph Camp, and Nuel Belnap. 1975. "A Prosentential Theory of Truth." In Grover 1992: 70-120.

Horwich, Paul. 1993. "Gibbard's Theory of Norms." *Philosophy and Public Affairs* 22: 67-78.

___. 1998a. *Meaning*. Oxford: Oxford University Press.

___. 1998b. *Truth*. Oxford: Oxford University Press.

Hume, David. 2000. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.

Loewer, Barry. 1987. "From Information to Intentionality." *Synthese* 70: pp. 287-317.

Lynch, Michael P., ed. 2001. *The Nature of Truth*. Cambridge, Mass.: MIT Press.

Moore, G.E. 1903. *Principia Ethica*. Cambridge, U.K.: Cambridge University Press.

O'Leary-Hawthorne, John, and Huw Price. 1996. "How to Stand up for Non-Cognitivists." *The Australasian Journal of Philosophy* 74: 275-92.

Quine, W.V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

___. 1970. *Philosophy of Logic*. Cambridge, Mass.: Harvard University Press.

___. 1992. *Pursuit of Truth*. Revised Edition. Cambridge, Mass.: Harvard University Press.

Railton, Peter. 1986. "Moral Realism" in Darwall, Gibbard and Railton 1997.

Schueler, G.F. "Modus Ponens and Moral Realism." *Ethics* 98: 492-500.

Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.