# The Yale Philosophy Review

*An Undergraduate Publication*

ISSUE 3 | 2007

**Issue III, 2007**

# Editors' Note

The third issue of the *Yale Philosophy Review* is different from its predecessors in many ways, some of which would likely go unnoticed without acknowledgement here. Most obvious is our new cover design. Less visible but noteworthy nonetheless is the inclusion of two rather lengthy interviews—with NYU's Ned Block and Robert Pippin of the University of Chicago—instead of one. We hope you will find them as interesting and even captivating as we did. Finally and perhaps most significantly for those who work on the *Review*, this is the first issue not produced under the guidance of its loving and beloved founders, James Martin and Amia Srinivasan '07. We miss them greatly and are humbled to follow after them.

Such changes suggest the alluring thought that the *Review* has moved into a new phase, but we nevertheless continue to feel the novelty of the project. It is a thrilling but uncomfortable position. What makes a journal something more than papers and computer data? It seems to have a life of its own, as though its editors were shepherds rather than creators, the adjuncts of something already real. Naturally, editors like to see themselves as these adjuncts, and we are no exception. We hope the future custodians of the *Review* will feel the same way for a long time to come, and wish them all the best of luck.

Special thanks are due to the Yale philosophy department—especially Professors Michael Della Rocca, Seyla Benhabib, and Kenneth Winkler—as well as the Yale Undergraduate Organizations Funding Committee and the John K. Castle Fund for their generous support, financial or otherwise. We also wish to thank the entire staff for their hard work throughout the editorial process. We hope you enjoy the issue!

Jordan Corwin & Yaron Luk-Zilberman
Editors-in-Chief

**THE Yale Philosophy Review**

Issue III, 2007



Email the editors at yalephilosophy@gmail.com.
For our online archive and submission guidelines,
visit <www.yale.edu/ypr>.

# Description in Goethe and Wittgenstein

Bihui Li
University of Chicago

Wittgenstein was strongly influenced in his formulation of the role of description in philosophy by Goethe's conception of description as a scientific method. However, despite retaining some superficial similarities, Wittgenstein's notion of the role of description in philosophy turns out to be an extreme morph of its scientific predecessor. Wittgenstein extends description's domain of application to such an extent that, unlike for Goethe, it becomes much more than a method for elucidating facts or principles, and conflicts with some of Goethe's original reasons for favoring description over explanation.

Goethe's scientific investigations and Wittgenstein's philosophical investigations mirror each other in their advocacy of description, as opposed to explanation, as the appropriate approach towards their respective intellectual endeavors. M. W. Rowe and Joachim Schulte, amongst others, have commented on this similarity, which is just one of the many ways in which Goethe influenced Wittgenstein. Given that Wittgenstein quotes Goethe's scientific writings approvingly at various points to support his own philosophical remarks, it is undeniable that Goethe's notion of scientific methodology did have a considerable influence on Wittgenstein. Despite this affinity in their patterns of thought, though, Wittgenstein's conception of philosophy makes a more radical point. While Goethe was proposing a new method for scientific investigation, he put it forward as a method that nonetheless worked towards traditional scientific objectives. Wittgenstein, on the other hand, is not merely trying to change how philosophy should attain its objectives, but to introduce new objectives altogether. Unlike Goethe, he is not just working within a predetermined set of objectives and suggesting a different method to attain those objectives. This difference between Goethe's and Wittgenstein's respective efforts to alter their fields of study can be most clearly seen in Goethe's arguments for the existence of archetypes of biological types. Although some aspects of Goethe's biology were adopted by Wittgenstein, these are but superficial similarities, for Goethe's archetypes have fundamental conflicts with Wittgenstein's later philosophy that go far beyond their similarities. In his notion of biological archetypes, Goethe makes a philosophical misstep that Wittgenstein is especially keen to avoid, a mistake that reveals profound differences in how they used description, and in why they both seized upon it as a solution to the problems they were concerned with.

It is probably in the Theory of Colours that Goethe's approach to science most strongly resembles Wittgenstein's approach towards philosophy. Goethe argues that his ob jective is to arrange his observations in a way that would allow the reader to judge the observations themselves and draw meaningful conclusions, sticking to descriptions and avoiding unnecessary theorizing. While Goethe concedes that "in every attentive look on nature we already theorise", he approached the phenomenon of color with the intent of not constructing any new theoretical structure in addition to the theoretical knowledge required to merely arrange the observed phenomena in a way amenable to further theoretical judgments (Theory of Colours, xx). The metaphor he uses to describe this aim is that of demolishing an old castle and rearranging the raw materials that had composed it into a fashion amenable for the construction of a new building. The old castle is Goethe's metaphor for Newton's theory of colors. Newton, he thinks, has been led astray in his study of colors by his obsession with fitting unnecessary theoretical baggage onto a few poorly chosen, secondary natural phenomena. Goethe instead favors offering an organized description of observed phenomena from which others can draw their own conclusions.

This approach towards science resembles much of what Wittgenstein would write about philosophy more than a century later. It parallels the Wittgensteinian theme, almost mantra, of not looking behind the phenomenon unnecessarily; of not postulating theoretical entities when the phenomena themselves are simpler and more direct presentations of what happens. This notion that any attempt to look behind the phenomena themselves is misguided occurs throughout Wittgenstein's later philosophy. Time and time again, Wittgenstein warns against the temptation of postulating complicated theories involving inner states of mind in order to explain everyday phenomena like our possession of beliefs, our seeing of aspects in a picture, and many others. For example, in criticizing the idea that the feeling of reading can be analyzed into various experiences that accompany the activity of reading, Wittgenstein points out that we are actually reading and not thinking about how we are reading; we have no experiences of letter and sound forming a unity, or of hearing the sound in the written word, or of any number of accompanying phenomena normally postulated to "explain" what the experience of reading consists in. If we invoke these supposed accompanying experiences, therefore, we are not contributing to our understanding what our normal experience of reading is. Thus we should avoid the temptation of trying to explain things by delving into unnecessary and irrelevant details:

> Nor should we be misled by the picture which suggests that these phenomena came in sight "on closer inspection". If I am supposed to describe how an object looks from far off, I don't make the description more accurate by saying what can be noticed about the object on closer inspection. (PI §171)

This can hardly fail to remind us of Goethe's criticisms of Newtonian optics as being fixated on details which are secondary to the phenomena in question. At one point, Wittgenstein even quotes the following remark of Goethe's: "Don't look for anything behind the phenomena; they themselves are the theory." (RPP, I:889) This was cited in the context of examining the question of what it means to have a sensation, to have a thought, to have a visual impression, and so on. As before, Wittgenstein is making the point that the enterprise of explaining thoughts, sensations, beliefs and so on by postulating mental states "behind" them is fundamentally flawed. It tries to do too much where a more direct strategy of simply describing the phenomena suffices. And, like Goethe's plan in the Theory of Colours, Wittgenstein's approach to philosophy is to describe things correctly and iron out confusions, so that what formerly seemed like philosophical problems turn out not to be problems at all:

> Philosophy simply puts everything before us, and

> neither explains nor deduces anything. —Since everything lies open to view there is nothing to explain. For what is hidden, for example, is of no interest to us. (PI §126)

Once again, we can see shades of Goethe's plan to lay his observations of colors out in an arrangement without delving into the kind of theoretical intricacies Newtonian scientists were involved in. Like Wittgenstein, Goethe thinks of his task as cleaning up confusions – in Goethe's case, these would be confusions in the theory of colors created by the explanations introduced by Newtonian optics. He regards it as an advantage of his method that it is unencumbered by any explanations, since this will allow his readers to "combine them according to his own judgment" (Goethe, Theory of Colours, xxvi). Presumably, said judgment, given a clear, organized set of studies as Goethe thought he had, would be able to see the value in those studies without significant difficulty, the same way Wittgenstein hoped that philosophical confusions laid bare can be clearly seen to be non-issues, without further confounding and contrived explanations.

Because Goethe is reacting directly to Newtonian science in his theory of colors, his criticisms of conventional explanatory approaches to science are emphasized the most in that context. It is also in the Theory of Colours that Goethe most scrupulously avoids imposing an overarching theoretical structure onto his observations, content with applying enough theoretical knowledge to organize the phenomena in a clear way without explaining them. Goethe's approach towards biology, on the other hand, at the very least edges towards an underlying explanation for the various forms of organisms we observe. Goethe claims, in explaining his approach towards biology, to use many of the same description-oriented principles that underlie his approach to colors, and indeed, these are not entirely spurious claims. However, I will argue that the reasoning which led him to conclude that archetypes of biological types must exist reveals a profound difference in his rationale for a descriptive science and Wittgenstein's rationale for a philosophy concerned only with description.

Although Goethe does not harp on it as much as he does in the Theory of Colours, the notion of science as a more descriptive than explanatory enterprise is also present in Goethe's approach to biology. In one essay, Goethe writes that the intention of morphology "is to portray rather than explain" ("Observation on Morphology in General", in Scientific Studies, 57). Fittingly, in his descriptions of the forms of plants in works like The Metamorphosis of Plants, Goethe never attempts to give explanations for the forms of plants. Instead, he describes them in an orderly fashion, proceeding, for example, from one part of the plant to another, as though writing entries in an atlas. There is, however, one aspect of his biological studies which distinguishes them qualitatively from his studies on colors. Goethe thought that for each biological type, there was an archetype which determined all the possible forms that the examples of the type could take. This archetype would represent all the members of the type and also represent the ideal of the type. Since it would also have to represent all stages of development and all possible environmental variants of its members, it cannot be a normal, static sensory image, and indeed Goethe writes in On Morphology that the archetype of all plants, the Urpflanze, appeared to him in the sensory form of a supersensory archetypical plant. Although it is supersensory in its representation of phenomena across time and space that cannot be captured in a single static image, it is also still "like" a picture in a sense. Goethe does try to draw various archetypes, and, in writing about his meeting with Schiller where he drew the Urpflanze for Schiller, referred to the Urpflanze as something he can see with his own eyes ("Fortunate Encounter", in Scientific Studies, 20). That he thinks it makes sense to draw the archetype suggests that the archetype is at least a partially sensible entity. Thus an archetype of a biological type can be thought of as something which is supersensory in its entirety, but with enough sensible characteristics that it would not be a completely meaningless enterprise to attempt to depict it as an image.

Thus far, it might seem that the notion of biological archetypes is in tune with Goethe's general approach towards science. The Urpflanze could be regarded as something like an overall description of all possible plants. After all, it does not attempt to explain why all plants should take the particular range of forms it circumscribes. It only dictates the boundaries and possibilities. However, when one considers the reasons Goethe gives for thinking that biological types must have archetypes, a deep-seated tension arises between the notion of the archetype and the supposedly primarily descriptive scientific methodology Goethe champions.

The most direct statement of Goethe's justification for the existence of archetypes occurs in the Italian Journey, where he writes: “There must certainly be [an Urpflanze ]. Otherwise, how could I recognise that this or that form was a plant if all were not built upon the same basic model?” (251) His postulation of the existence of the Urpflanze, then, is based on the assertion that there must be an Urpflanze if we are able to recognize any forms as plants. Goethe does not explain how the existence of an Urpflanze would enable us to recognize disparate forms as plants, but presumably it works something like this: we compare, unconsciously or consciously, forms of things we observe to the Urpflanze, and thereby determine if the things we observe are plants. Whatever the exact mechanism is, the implication is that we possess some notion of what the Urpflanze is like, even if we cannot describe this notion explicitly, and we use our incomplete, implicit knowledge of the Urpflanze to determine if the forms we observe are plants. That Goethe takes on the task of collecting enough observations of plants to help him better formulate the appearance of the Urpflanze shows that he too thinks that although we know enough of the Urpflanze to be able to recognise plants, we do not have a complete conception of what it is.

This, I will argue, is where Goethe commits the mistake he repeatedly marks as a temptation to be avoided in scientific inquiry, and lets his poetic imagination distract him from the proper, descriptive scientific enterprise as defined by himself. It is reasonable to conclude, from the fact that we are able to recognize certain things as plants, that we have a notion, however vague, of what any plant should look like. However, to proceed from there to the conclusion that there exists an archetype of all plants, let alone an archetype with the various improbable attributes he attaches to it — partially sensible, ideal, and supersensory, smacks of a move that Goethe highlights as a common temptation to scientists when he writes “Thinking man has a strange trait: when faced with an unsolved problem he lies to concoct a fantastic mental image.” (“Selections from Maxims and Reflections,” in Scientific Studies, 308) In concluding that an archetype must exist, Goethe has made just the kind of hasty conclusion fueled by imagination that he warned his peers to avoid. Perhaps this does not seem terribly significant. One cannot reasonably expect Goethe’s broad spectrum of intellectual activity to be completely free of internal inconsistencies. I will argue, however, that this inconsistency has a profound significance, for it points to an irreconcilable difference between the motivations behind Goethe’s method of science and Wittgenstein's notion of purely descriptive philosophy. It is a difference which, on a philosophical level, renders the similarities previously explicated quite superficial.

Goethe's justification for the existence of archetypes is a mistake in philosophy that Wittgenstein, in his descriptive notion of philosophy, takes especial pains to avoid. One could even plausibly argue that the later Wittgenstein's notion of a purely descriptive, confusion-dispelling philosophy was developed in reaction to intellectual leaps of this sort — that is, intellectual leaps from the occurrence of types of human behavior that one would regard as simple in an everyday sense to the existence of less realistically grounded entities that supposedly underlie or explain the observed phenomena.

The avoidance of such intellectual leaps is a recurring theme in Wittgenstein's later philosophy. He particularly warns against the temptation of concluding, from the fact that we speak as though our experiences of phenomena like pain, aspect perception and doubt are accompanied by images, that there are pictures in our minds accompanying these experiences. In his discussion of pain, he writes that "[t]he image of pain is not a picture and this image is not replaceable in the language-game by anything we should call a picture." (PI §300) The temptation is to think that since we ordinarily point to a picture of the object we perceive if asked what our mental image is, mental images are like pictures. Yet, Wittgenstein argues, our habit of pointing to the perceived picture in response to the question "What is the mental image you have now?" is the only connection between mental images and pictures, and we have no reason to think that because of this connection, the two are similar in other respects. This is what he means when he writes:

> But if we ask, "What is the difference between a visual picture and an image-picture?" — the answer could be: The same description can represent both what I see and what I imagine.(RPP II §69)

Since it is only because we tend to describe sensory impressions and mental images in the same way that we think mental images could be pictures in our minds, he thinks that "[t]he tie-up between imaging and seeing is close; but there is no similarity"(RPP II §70). To him, seeing and forming images are fundamentally different phenomena. We cannot, he says, even think of them as different activities, for "that is as if one were to say that moving and losing in chess were different activities." (RPP II §138) Activities are at least phenomena of the same type; they can be meaningfully compared to one another. We can compare running and swimming, but we cannot compare running and winning, for winning is not even an activity. The relationship between seeing and forming mental images is, to Wittgenstein, more like the relationship between running and winning than the relationship between running and swimming. Forming mental images does not stand to seeing the way running stands to swimming, so we must treat the formation of mental images differently from how we treat the seeing of pictures.

Wittgenstein makes a similar argument in his discussion of aspect perception. Here, he uses the metaphor of different dimensions to distinguish between the sensory perception of a picture containing certain aspects and the mental image of those aspects. The sensory perception of a Necker cube, we would like to say, is just the arrangement of lines constituting the raw picture itself. The mental image is something like an image of a cube in one orientation or another. They are clearly different, yet we would point to the same picture whether we are asked to show what the ob ject of our sensory perception is or toshow what the aspect we are perceiving is. This behaviorial artifact tempts us to say that "the aspect is as much 'seen' as are the shape and colour" (RPP I §1024). But far from suggesting a similarity between the perceived aspect and the picture itself, Wittgenstein argues that the fact that we have one and the same answer to two profoundly different questions shows that there is no room to accommodate the notion of the perceived aspect as a picture. The possible logical space of "picture", as such, is already occupied by the raw arrangement of lines itself. We know that the perceived aspect and the picture cannot be the same thing, so we must find room for it in another dimension of logical space. To Wittgenstein, the perceived aspect cannot even be another type of picture, but something of a fundamentally different nature. The analogy he uses (PI, II §xi, 200-201) is that imaginary numbers are fundamentally different from real numbers even if they are connected to real numbers by the square root operation, and this difference is manifested in how there is no room for imaginary numbers along the dimension of the real number line, just like how there is no room in our conceptual apparatus for treating mental images like another type of picture without abusing the very notion of a mental image. Imaginarynumbers are not just another type of real number, and it is for a similar reason that the mental images we have when perceiving aspects are not just another type of picture.

Wittgenstein would regard Goethe's argument for the existence of archetypes as an example of the misuse of pictures in our analogies about how we think. In a way, Goethe goes even further than Wittgenstein's targets in the critiques explained above, for Goethe's pictorialization of our ability to recognize natural forms is more explicitly pictorial than the theories that postulate pictures in our minds. A typical proponent of inner mental pictures might say that we see the inner picture with our mind's eye. Goethe, however, assigns a greater measure of reality to the archetype, for it is not just something that he perceives with his mind's eye, but, as he said to Schiller, something that he can see with his physical eyes. What Goethe has done, then, is to explain the "feeling" of recognition we get when we see a plant and realize that it is a plant by introducing a kind of picture of a plant. Unlike Wittgenstein's idea of what mental images are in relation to ordinary pictures, the archetype is a solution that is along the same dimension as ordinary pictures. It may have special qualities that ordinary pictures do not, but its nature is still that of a picture, and indeed it is its pictorial nature that makes it so attractive a notion. Its pictorial nature allows it to be drawn, and hence allows us to apprehend it with our senses as directly as we would any other picture, to see all possible plants before our eyes. Its pictorial nature gives it an immediacy and vivacity that abstract theories lack. It is the most important respect in which the Urpflanze has an edge over the kind of abstract theory, in the tradition of Newton's theory of colors, that Goethe would have objected to. It is also the respect in which it is fundamentally antithetical to Wittgenstein's philosophy. Not only does Goethe employ a tactic that Wittgenstein is reacting against, that of transferring the idea of a picture into realms where it does not belong, he also amplifies the idea of a picture into something like a super-picture and treats this as the basis for our recognition of biological types.

This has a significant bearing on Goethe's conception of the scientific method and his justification for the existence of archetypes. Goethe's rationale for a descriptive rather than explanatory science is offered as an alternative to Newtonian methodology, but it is an alternative that is along the same dimension as Newtonian science. If we think of Newtonian science as a real number, Goethe's method would be another, different, real number, and both would lie on the same, unitary real number line. Wittgenstein's contribution to philosophy, however, is akin to inventing imaginary numbers, along a different dimension to the existing philosophical theories along the real number line. Wittgenstein opens a new dimension and hence avoids having to play the same game that previous philosophers played. In other words, Goethe's scientific methodology is not a radically different concept of science in the way that Wittgenstein's descriptive concept of philosophy is a radically different concept of philosophy. It departs from Newtonian science in the methodology it recommends for attaining scientific truth, but it is a methodology that still pushes science in the same general direction as before. Although Goethe thinks that scientific truth is to be derived primarily through description rather than explanation, he does not give up the idea of underlying principles behind nature. Indeed, Goethe sometimes writes in hope of a time when "we will discover the true underlying principles of matter" ("Excerpt from Toward a Theory of Weather ", in Scientific Studies, 150). In contrast, Wittgenstein is positively allergic to all attempts at discovering underlying principles in philosophy. This is why Goethe finds the notion of archetypes such a useful and plausible idea, while Wittgenstein would have the opposite attitude. The archetype, for Goethe, is a kind of underlying principle. Even if it is not explanatory in the way Newtonian mechanics is, it is still something behind all the appearances that a particular biological type can assume. And description, for Goethe, is the method by which one can discover the complete form of the archetype. Thus for Goethe, unlike for Wittgenstein, description is not the end of the road. Description is instead a tool for finding the underlying truth.

In short, Wittgenstein has a more deep-cutting criticism of philosophy, and more revolutionary recommendations to offer for philosophy, than Goethe has for science. Wittgenstein sought to show that philosophy should not be a matter of directional progress in the first place, while Goethe sought to show that his descriptive method of doing science should be employed to move science further along in its directional progress. Goethe worked within the existing framework and suggested description as a superior alternative to Newtonian theoretical explanations, without denying the ultimate objective of science, which was to discover the underlying principles of nature. Wittgenstein, too, champions description over explanation, but for the deeper reason that attempts to explain the phenomena he was concerned with are, to borrow a turn of phrase from the physicist Wolfgang Pauli, not even wrong, said attempts being based on such a deeply mistaken view of philosophy that they cannot even be considered as candidates for an approach towards philosophy. In contrast, Goethe thought that Newtonian science was wrong; that it was taking the wrong direction, but within a context where it at least had the possibility of being right or wrong. In drawing inspiration from Goethean description to formulate his alternative to traditional philosophy, Wittgenstein effectively appropriated what was meant to be a mere tool for other objectives and elevated it to be the only objective. It is as though he had cultivated, from a seed of Goethe's thought, a completely alien plant, one that certainly did not fit any preexisting archetype.

## Works Cited

Goethe, Johann W. v. *Italian Journey*. Trans. W. H. Auden and E. Mayer. San Francisco: Pantheon Books, 1962.

———. *Scientific Studies*. Trans. Douglas Miller. New York: Suhrkamp Publishers, 1988.

———. *Theory of Colours*. Trans. Charles L. Eastlake. London: John Murray, 1840.

Rowe, M. W. "Goethe and Wittgenstein." *Philosophy* 66 (Jul. 1991): pp. 283-303.

Schulte, Joachim. "Goethe and Wittgenstein on Morphology". In *Goethe and Wittgenstein: Seeing the World's Unity in its Variety*, ed. Fritz Breithaupt, Richard Raatzsch, and Bettina Kremberg. Frankfurt: Peter Lang, 2003.

Wittgenstein, Ludwig. *Philosophical Investigations*. Trans. G. E. M. Anscombe. Oxford: Blackwell Publishers, 1953.

———. *Remarks on the Philosophy of Psychology*. 2 vols. Trans. G. E. M. Anscombe. Chicago: University of Chicago Press, 1980.

# Are We Bound to Uphold Rawlsian Justice?

Alice Evans
University of Nottingham

*A Theory of Justice* maintains that we are morally bound to further those institutional arrangements that support those principles that would have been agreed to by contracting parties in the original position. However, some critics have rejected the implicit premise that hypothetical contracts yield contractual obligations. But this critique is misplaced according to a different interpretation of the contract's role. Rawls arguably claims that justice is binding and that in virtue of specifying the content of justice the hypothetical contract is likewise binding. To determine whether we are bound to uphold Rawlsian justice, I shall discuss both approaches and then further analyse charges of triviality, circularity, an alleged similarity to intuitionism and the contractarian rebuttal of utilitarianism.

## The Contractual Interpretation

Let us first consider the *contractual* interpretation. Perhaps we are contractually bound to uphold the terms of Rawls's hypothetical contract because partially veiled contractors would have agreed to its terms in the original position. But Dworkin, already presupposing that Rawls invokes a contractual obligation, denies that hypothetical contracts are relevantly similar to actual contracts. Dworkin argues that 'my hypothetical agreement does not count as a reason, independent of … other reasons, for enforcing … rules against me, as my actual agreement would have[1]'. Additionally, even if hypothetical contracts were relevantly analogous to the actual version, it may be further disputed that either can give rise to obligations. A contract (actual or hypothetical) does not logically entail that parties are morally bound to adhere to the respective terms. A contractarian solution might be to postulate a prior contract whereby parties contract that the terms of future contracts are binding. But this reply unleashes an infinite regress, incapable of yielding the

[1]Dworkin, R. "Justice and Rights" in *Taking Rights Seriously*. London: Gerald Duckworth & Co. Ltd., 1981. p. 151

necessary foundations for contractual obligation. A further option would be to argue that, as a necessary truth, contracts do give rise to obligations. But positing this premise, in defence of Rawls, seems ironic, given that contractarianism is offered as a solution to the apparent unavailability of such truths in ethics[1]. So that which is required for contracts to give rise to obligations appears inconsistent with the assumptions of Rawls's contractarianism.

## The Moral Interpretation

A solution to this problem might be to deny that we are bound to uphold the terms of the hypothetical contract in virtue of contractual obligation. There is substantial textual support for the alternative. Rawls maintains that we have a natural duty of justice to uphold whatever is agreed upon in the original position because that situation 'best expresses the conditions that are widely thought reasonable to impose on the choice of principles'[2]. In virtue of these principles expressing the content of justice, we are morally bound to uphold their terms, or so Rawls would claim.

This interpretation evades the aforementioned criticism of the contractarian nature of Rawls's work[3] because moral obligation does not require that contracts can, of themselves, give rise to obligations. Dworkin's contention that hypothetical contracts are not contractually binding thus misses the mark, for he appears to have misinterpreted moral obligation for contractual obligation. It is not relevant that we did not actually consent to these principles. The crucial question is whether there is sufficient moral argument for the original position and the formulation of the ensuing terms of the hypothetical contract.

Given that the moral interpretation appears most plausibly Rawlsian and most defensible, I shall now investigate whether the original position lays sufficiently moral foundations so as to determine whether Rawls can secure the required premise

[2]Rawls, J., *A Theory of Justice*. Oxford: Oxford University Press, 1999. p. 19

[2]Ibid. pp. 21, 17, 37, 99, 334

[3]Such as but not limited to Robins, M.H. "Promissory Obligations and Rawls's Contractarianism" in *Analysis,* Vol. 36, No. 4. (June 1976), p. 195; Bloom, A. "Justice: John Rawls vs. the Tradition of Political Philosophy" in *The American Political Science Review* 69 (1975), pp. 648-662; Schaefer, D.H. *Justice of Tyranny? A Critique of John Rawls's "Theory of Justice"*. Port Washington: Kennikat Press, 1979. pp.39-41; Stark, C.A. "Hypothetical Consent and Justification" in *The Journal of Philosophy,* Vol. 97, No. 6. (June 2000), p.321

that whatever is agreed upon by contracting parties specifies the content of justice and is thus morally binding.

## The Case for the Moral Appropriateness of the Original Position

Rawls argues that we should imagine ourselves abstracted from those facts about ourselves which enable bias and from there determine what might be rationally agreed to by mutually disinterested parties[1]. By formulating the original position so that no one yields greater bargaining advantages than others[2], Rawls ensures that the principles agreed upon by the contracting parties secure fairness. Rawls argues that equal consideration is owed to all persons[3], as Kantian 'ends in themselves'[4], in virtue of their capacity to have a sense of their good and of justice (i.e. their 'moral capacity')[5].

Let us now analyse the above-mentioned claims - which indirectly mould the resulting principles, by shaping the circumstances of the hypothetical contract. There are several reasons why it is crucial that 'the principles of justice do not depend upon strong assumptions'[6], such as a substantive theory of the good. Given Rawls's commitments, firstly to reasonable pluralism[7], secondly to ensure unanimous agreement[8] and thirdly to avoid charges of triviality and circularity. The original position must not be constructed in such a way that it constrains contractors into choosing an unjustifiably non-ethically-neutral set of principles. But Rawls also admits that 'a political conception must draw upon various ideas of the good… The question is: subject to what restriction may political liberalism do so?'[9]. Rawls decides that the ideas included must be political ideas, those shareable by citizens regarded as free and equal so that 'accepting the political conception does not

[1]Rawls, J., 1999. p. 129

[2]Ibid.p. 121; Rawls, J. *Justice as Fairness*. Massachusetts: Harvard University Press, 2001.p. 235

[3]Rawls, J., 1999. p. 475, 511

[4]Ibid. p. 442

[5]Ibid.pp. 17, 115-116, 289, 442-47; Rawls, J. *Justice as Fairness*. p. 233

[6]Rawls,J., 1999. p. 111, 512

[7]Ibid. p. 347, 393

[8]Rawls, J. *A Theory of Justice*. Oxford: Clarendon Press, 1972. pp. 520-521

[9]Rawls, J. "The Priority of Right and Ideas of the Good" in *Philosophy and Public Affairs*, Vol. 17, No. 4 (Autumn 1988), p. 264

not presuppose accepting any particular comprehensive religious, philosophical or moral doctrine'[1].

Rawls maintains that moral equality is a non-particular thin theory of the good[2]. But it seems that any independent reasons that might be invoked for defending moral equality, as based on his normative commitment to Kantian self-determination and personhood, will inevitably invoke some substantive concept of the good. Another option would be to argue that moral equality is so fundamental that it is not just one value amongst many but has morally privileged status. But again, this latter claim is precisely what Rawls seeks to (and indeed must) avoid[3]. He cannot extol the value of equality and build his original position upon this premise whilst remaining neutral between all competing conceptions of the good. It is, therefore, not the case that 'the original position is used to determine the content of justice [and] the principles which define it'[4], for a theory of the good has already been built into the original position.

## Reasonable People and Triviality

Rawlsians might concede that the original position is not predicated upon ethical neutrality[5] but defend such stipulations in virtue of their being widely-held intuitive assumptions. Rawls likewise argues that such predicates are necessary to ensure certain ethical constraints[6]. Justice as fairness is also said to achieve the stated goal of stability[7] because it is supportable by every 'reasonable person'[8], regardless of their respective reasonable comprehensive views. Such support would be forthcoming because reasonable persons are defined, by Rawls, as recognising moral equality and the plurality of reasonable comprehensive doctrines. Such doctrines are formulated by theoretical reason and practical reason, so that they

[1]Rawls, J., 1988. p. 252
[2]Ibid. p. 289
[3]Rawls, J., 1999. pp. 512, 587
[4]Ibid. p. 511
[5]Ibid. p. 507
[6]Ibid. p. 512
[7]Ibid. p. 6
[8]Ibid. p. 12

are: internally coherent, able to balance conflicting values and flexible to revision[1]. Reasonable persons will thus 'think it unreasonable to use political power… to repress comprehensive views that are not unreasonable, though different from their own'[2], and will thus support justice as fairness.

Rawls's justifications thus seem somewhat trivial. He stipulates that a concept of justice must be agreeable to reasonable people but then defines reasonable people in a way that ensures that the only conception of justice that they could find agreeable is his own conception. Sadurski likewise comments,

> If the original position is designed in such a way as to exclude the possibility of choice of some principles contrary to prior moral judgments, then the parties to the original contract are nothing more than puppets moved by the invisible hand of the author and are bound by his moral constraints' which 'pushes the matter of pre-contractual, substantive moral intuitions from the position of the contractors to the position of their 'creator'.[3]

## A Rawlsian Reply

To maintain that a conception of justice should be acceptable to reasonable people and rebut the abovementioned such charges of triviality, Rawls must allow for the conceptual possibility of reasonable people denying that his hypothetical contract is an agreeable way of respecting moral equality[4]. Such disagreements are surely plausible given that Rawls maintains that the form of equality is left unspecified by his thin theory of the good[5].

But this reply prompts further problems. If the original position is only justified by an appeal to its agreeableness to reasonable persons but some such persons do not regard the original position as agreeable, the original position would then appear

[1]Rawls, J. *Political Liberalism*. New York: Columbia University Press, 1995. p. 59

[2]Ibid. p. 60

[3]Sadurski, W., "Contractarianism and Intuition" in *Australasian Journal of Philosophy,* Vol. 61, No. 3 (September 1983).p. 241

[4]Rawls, J., 1999. p. 12

[5]Ibid. p. 511

unjustified. Indeed, there appears to be a vast range of reasonable people who do not find the original position agreeable. Other theorists have interpreted respect for equality as requiring Libertarianism; treating persons as ends in themselves so as to respect their self-ownership. Other interpretations have included resource and welfare egalitarianism. A Utilitarian might also maintain that by denying privileged claims to goods, subjects are treated as moral equals. Even those that grant that moral equality necessitates justice as fairness might still deny that the original position ensures such fairness and hence dispute its agreeableness As Dworkin comments, 'the fact that a particular choice is in my interest at a particular time, under conditions of great uncertainty, is not a good argument for the fairness of enforcing that choice against me under later conditions of much greater knowledge'[1].

Although such interpretations do not prove that the content of justice is not specified by the terms of the hypothetical contract, they do suggest that not all reasonable persons will regard the original position as agreeable. If Rawls's only defence of the original position is that it 'embodies widely accepted and reasonable constraints on the choice of principles'[2] but we also discover that the original position is not universally regarded as reasonable then, to defend his position, Rawls must show that dissenting opinions are unreasonable. Rawls might argue that the Libertarian alternative is 'unreasonable' in virtue of 'permit[ting] distributive shares to be improperly influenced by factors so arbitrary from a moral point of view'[3]. He might likewise disallow the 'unreasonable' opinions of strict egalitarians. Indeed, Rawls regarded envy as an irrational motivation for those in the original position[4]; believing that it would only lead to 'collective disadvantage', 'mak[ing] everyone worse off'[5]. Rawls might further discount Dworkin's retort, of the original position being unfair; for if people knew the facts about their actual circumstances that might ensure unfairly unequal bargaining power. If all those who disagree with Rawls's claim that the terms of the original position specify the content of justice are 'unreasonable', Rawlsian justice is thereby agreeable to all reasonable people. So, in virtue of being agreeable to all reasonable people, the original position is

[1]Dworkin, R., p.153
[2]Rawls, J., 1999. p. 12
[3]Ibid. p. 63
[4]Rawls, J., 1999. p. 124
[5]Ibid. p. 124

morally relevant and that the ensuing terms of the hypothetical contract thus specify the content of justice. If justice is binding, we thereby are morally bound by the terms of the hypothetical contract.

For if Rawls can only defend his position by claiming that the persons that he regards as reasonable will find it agreeable, and no independent account of reasonableness is given, his account seems trivial. If Rawls cannot appeal to an independent account of reasonableness, then that justification for the original position (i.e. that it is reasonable to reasonable people) cannot be defended non-trivially. Rawlsians might reply that the theory's justification rests upon how well it coheres within itself and with our intuitions[1]. But in truth the theory's justification rests on how well it coheres with particular people's intuitions. In short that he regards it as agreeable. As Nozick comments, 'Rawls is merely repeating that it seems reasonable; hardly a convincing reply to anyone to whom it doesn't seem reasonable'[2].

## Intuitionism

The hypothetical contract is intended to act as an independent justificatory device, supplementing reflective equilibrium's direct appeals to intuitions[3]. But the contract can provide no such role for it has already been tailored so that it, in itself, fits with and, in turn, generates principles that cohere with Rawls's intuitions[4]. Although Rawls distinguishes himself from intuitionists by providing a constructive answer for adjudicating competing claims, given that his answer is itself a product of his intuitions then, just like intuitionism, 'there exists no higher-order constructive criteria for determining the proper emphasis for the competing principles of justice'[5]. With no independent justificatory force, beyond intuitive appeal, Rawls's hypothetical contract 'is merely an expository or didactic device for explaining principles already held'[6].

---

[1]Ibid. pp. 19, 105
[2]Nozick, R. *Anarchy, State and Utopia*. Blackwell Publishing: Oxford, 2003. p.196
[3]Rawls, J., 1999. p. 37 (21)
[4]Ibid. p. 125
[5]Rawls, J., 1999. p. 30
[6]Sadurski, W., 1983. p. 238

The aforementioned lack of foundations not only threatens the universalist intentions implicit in A Theory of Justice[1] but also threatens the validity of those principles later chosen. In cases where our considered moral judgments conflict with the principles derived from the original position, coherence requires that one must be adjusted to fit the other. But if either change ensures coherence and we have no guidelines, beyond intuitions, for adjustment decisions, then there is no answer as to whether we should change the judgment so that it fits the principles derived from the original position(A) or change the circumstances of the original position so as to develop principles which fit the judgement (B). But this justification appears like the very intuitionism that Rawls finds so problematic; for in this case it is simply a matter of 'strik[ing] a balance by intuition, by what seems to us most clearly right'[2]. Considerations of coherence tell us to chose, between (A) and (B), but they cannot tell us whether to chose (A) or (B). Short of any further criteria by which to adjudicate, any adjustment decision will be intuitionist.

For instance, Rawls argues that when intuitions conflict with those principles chosen in the original position (such as over economic reward, desert, contribution to society and the free-market[3]) we should modify or abandon those intuitions. But if the original position is also tailored to fit certain intuitions, this stipulation merely implies that one particular cluster of intuitions trumps all others. Without any independent justification for the primacy of one cluster, such trumping appears arbitrary.

Rawlsians might reply that the intuitions that justify the original position are significantly different from all others. Perhaps the former kind is particularly fundamental and sufficiently broad so as to elevate the original position, thereby enabling 'a more general scheme for determining the balance of precepts'[4]. But intuitions do not appear so obviously divisible, least of all not universally and non-arbitrarily. But perhaps, as Rawls maintains, those intuitions that are held more intensely, such as regarding religious intolerance[5], should be kept, whilst those which

[1]Rawls, J., 1999. p. 114

[2]Rawls, J., 1972. p. 34

[3]Rawls, J., 1972: On economic reward and moral desert, pp.103– 4, 88– 9, 310– 15, 273– 77; on the individual's contribution to society, pp. 307– 8/ 270– 1. (See pp. 19-20, 206)

[4]Ibid. p. 31

[5]See pp. 19-20, 206

are held less intensely, such as regarding wealth distribution, should be modified in accordance with the dictates of the original position. But such subjective relative confidence could easily be different for different reasonable persons. In addition, this method of adjudication seems much like the intuitionism that Rawls repeatedly repudiates. Given that Rawls regards the 'assignment of weights [as] an essential part of a conception of justice' and that his own theory fails this pre-requisite, Rawlsian justice seems to fail, by the author's own standards. As Hare comments, 'Rawlsian justice is arguably more systematic than [pluralist intuitionism], but no more firmly grounded'. Rawls merely 'intuits the validity of a single method and erects his entire structure of moral thought on this'[1].

## The Contractarian Rebuttal of Utilitarianism

In order for us to be morally bound to uphold Rawlsian justice, in virtue of the terms of the hypothetical contract specifying the content of justice, those terms must specify Rawlsian justice. Indeed, Rawls states that his contractarian argument rebuts all forms of utilitarianism.

Keen to ensure that the principles of justice are chosen impartially, Rawls stipulates that contractors have no knowledge of the probabilities of their unveiled selves being in anyone's position[2]. But impartiality would also be obtained if contractors thought that they had an equal chance of being in anyone's position[3]. But, as Rawls admits, 'this idea of the initial situation leads naturally to the average [Utilitarian] principle'. This is because under such conditions of risk, one's 'prospects are identical with the average utility for each society'[4]. Hence it would be rational for contractors to choose principles whose acceptance would maximise the average level of well-being. Given this professed entailment, Rawls must defend his stipulation of uncertainty, rather than risk, so as to justify the claim that the hypothetical contractors will chose non-utilitarian principles. He could appeal to his earlier claim that certain formula can be rejected if their implications are too

[1]Hare, R. M. "Rawls's Theory of Justice—I" in *The Philosophical Quarterly,* Vol. 23, No. 91. (April 1973), p.146

[2]Rawls, J. 1999. p. 149

[3]Parfit, D. "What We Could Rationally Will" in *The Tanner Lectures on Human Values (*2002). p. 345

[4]Rawls, J., 1999. pp. 165-66.

counter-intuitive[1]. Our non-utilitarian beliefs might then justify the stipulation of uncertainty, rather than risk. But this argument would not then suffice for the Rawlsian claim that the principles then decided upon, under conditions of uncertainty, justify our non-utilitarian beliefs. Having justified particular principles by an appeal to particular beliefs, those beliefs cannot then be justified by an appeal to those same principles.

## Five problems

Returning to the question of whether we are morally bound to uphold Rawlsian justice, there seem to be five problems; cumulatively lessening the plausibility of the required premise that the terms of Rawls's hypothetical contract specify the content of justice. Firstly, Rawls's justifies his conception by appealing to intuitions but, he only counts those he deems 'reasonable'. Rawlsian justice is then trivially justified.

Second, given his ultimate appeal to intuitions, Rawls does not justify why one set of his intuitions trump another set. Failure to justify the priority of one set makes his account appear intuitionist. Thirdly, such implicit intuitionism makes Rawlsian justice, by his own standards, unsuited for a political conception of justice.

Fourthly, in failing to provide any independent justification for the original position, Rawls fails to show the moral relevance of the hypothetical contract and in turn fails to show it specifies the content of justice. Scanlon likewise comments that 'however internally coherent [Rawls's theory may render] our moral beliefs, [without any external vantage point] the nagging doubt may remain that there is nothing to them at all'[2]. In reply, Rawls's professed ambivalence over meta-ethical claims[3] suggests that his conception was not intended to be transcendental but rather sufficient to ensure a well-ordered society, bound to uphold certain fundamental principles.

But even if it is the case that universal acceptance of Rawlsian principles of justice leads to a well-ordered stable society, this does not suffice to show that we

---

[1]Ibid. p. 18

[2]Scanlon, T. M. "Contractualism and Utilitarianism," in Sen. A. , and Williams, B. *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 1982. p. 106.

[3]Rawls, J., 1999. p. 19

are bound. If Rawls cannot show the moral relevance of the original position then he will not have shown that the terms of the hypothetical contract specify the content of justice and hence will not have shown that such terms are morally binding.

Fifthly, even if we are bound to uphold whatever terms are specified by the hypothetical contract it is not clear that those terms are Rawlsian rather than average utilitarianism, given that Rawls's stipulation of uncertainty rather than risk appears arbitrary. Given that '[Rawls's] principles of justice have their justification in the fact that they would be chosen'[1] but it now transpires that such principles need not be chosen, the content of Rawlsian justice is thereby unjustified.

## Conclusion

I have argued that if we are bound by the hypothetical contract it is either in virtue of its valid contractual formation, making us contractually bound, or of its capacity to specify the content of justice, making us morally bound. The latter interpretation avoids the problem of how contracts might give rise to binding obligations but Rawls has not shown that the hypothetical contract is morally relevant. The charges of triviality, circularity, intuitionism and arbitrariness cumulatively make it implausible that the hypothetical contract specifies the principles of justice, even by Rawls's own standards of justification. Furthermore, even if the contract were morally binding, its terms need not be Rawlsian, rather than average utilitarianism. In sum, it seems implausible that anyone is bound to uphold Rawlsian justice.

[1]Rawls, J., 1999. p. 37

## Works Cited

Bloom, Allan. “Justice: John Rawls vs. the Tradition of Political Philosophy” in *The American Political Science Review* 69 (1975), pp. 648-662

Dworkin, Ronald. “Justice and Rights.” In *Taking Rights Seriously*. London: Gerald Duckworth & Co. Ltd., 1981.

Hare, R. M. “Rawls’s Theory of Justice—I” in *The Philosophical Quarterly*, Vol. 23, No. 91. (April 1973), pp.144-155.

Nozick, Robert. *Anarchy, State and Utopia*. Blackwell Publishing: Oxford, 2003.

Parfit, Derek. “What We Could Rationally Will.” In *The Tanner Lectures on Human Values* (2002).

Plato, Grube, G. M. A. (trans.) *The Republic*. London: Hackett Publishing, 1992.

Rawls, John. *Justice as Fairness*. Massachusetts: Harvard University Press, 2001.

———. *A Theory of Justice*. Oxford: Oxford University Press, 1999.

———. “The Priority of Right and Ideas of the Good.” In *Philosophy and Public Affairs*, Vol. 17, No. 4 (Autumn 1988), pp. 251-276.

———. *Political Liberalism*. New York: Columbia University Press, 1995.

———. “Kantian Constructivism in Moral Theory.” In *The Journal of Philosophy* 77 (1980).

Robins, M. H. “Promissory Obligations and Rawls’s Contractarianism.” In *Analysis*, Vol. 36, No. 4. (June 1976), p. 195.

Sadurski, Wojciech, “Contractarianism and Intuition.” In *Australasian Journal of Philosophy*, Vol. 61, No. 3 (September 1983) p. 231-247.

Scanlon, T. M. “Contractualism and Utilitarianism.” In Sen, Amartya and Williams, Bernard. *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 1982.

Schaefer, D. L. *Justice of Tyranny? A Critique of John Rawls’s “Theory of Justice.”* Port Washington: Kennikat Press, 1979.

Stark, C. A. “Hypothetical Consent and Justification.” In *The Journal of Philosophy*, Vol. 97, No. 6. (June 2000), pp. 313-334.

# Possibilism and Frege's Puzzle

Daniel Koffler
Yale University

> We can go to extremes of intensionality, if we like. Semantic rules can be stated entirely in terms of intensions, while extensions go unmentioned. But when we do, it seems for all the world as if we've gone purely extensional instead! Let me explain.
>
> —David Lewis, "'Tensions"

This is a paper for those of us who believe that the jury is still out on possibilism.[1] My aim is to explore an internal tension that any successful theory of possibilism is going to have to come to terms with. The tension is this: Possibilism involves a set-theoretic reduction of properties and propositions; since properties and propositions are intensional entities, whereas sets are extensional entities, possibilism is a species of extensional reductionism; but possibilists (owing to their rejection of the primitive trans-modal identities of direct reference theory) are also broadly Fregean in their semantics. Possibilists are therefore committed to explaining the phenomenon of intensionality, i.e. the failure of substitution of equivalents salva veritate in various contexts, as a genuine feature of linguistic meaning, and not something to be explained away by pragmatics. Thus it would seem that possibilists cut their legs out from underneath themselves, reducing away the very entities they need to make their semantics work. My purpose is to give an account of substitution failures consistent with possibilist extensional reduction.

Following the rhetorical conventions of possibilism, this account will be given in terms of possible worlds, possible individuals, and locations in modal space; whether the possibilia and realms of possibilia should be taken as real existents or as ersatz model-theoretic representations remains an open question[2] — i.e., I remain neutral about robust modal realism versus modal ersatzism (hereafter "MR" and "ME," respectively). Finally by way of introduction, though this is a paper

---

[1] Roughly speaking, I use "possibilism" to denote agreement with some version of the indexicality theory of actuality, rather than as the customary synonym for 'modal realism'. Hence, on my usage, possibilism encompasses modal realism and some versions of ersatzism; the point is to contrast such possibilism from philosophies which reject the indexical theory, no matter whether they believe that possible worlds exist as abstract objects or do not exist at all. The reason behind this usage will be explained in §1.

[2] As does the question of whether possibilism requires the positing of possible individuals in addition to possible worlds. Briefly, modal realism answers affirmatively while modal ersatzism answers negatively, so a decision in favor of one or the other should elliptically decide the answer to this question.

for possibilists, committed actualists are welcome to kibbitz the proceedings. Indeed, they might find it valuable to do so, as the weight of theoretical artifice (as they would see it) involved in the possibilist account of intensionality may serve to confirm the suspicion of actualists that possibilism is simply too implausible to be worth considering.

---

## 1. Defining Possibilism, Possibilist Reduction of Actuality

There are two general forms of possibilism. MR: All possible worlds are existing concrete objects, one of which happens to be actual.[1] ME: Possible worlds are abstract objects in the actual world, one of which accurately represents[2] the actual world, the rest of which do not. A possible world for a strong MR-ist is a big rock, the actual world one big rock among infinitely many.[3] Whereas under ME, worlds are "real" in the sense that they are abstract objects, and really exist only inasmuch as ME ontology admits of the existence of abstract objects. Or, from the perspective of ME, MR makes little sense because the world is not "a way things could have been"; rather, one of the ways things could have been corresponds to the way things are.

Robert Stalnaker, an able defender of one kind of ME, takes a position that we might call in re modal realism (as opposed to Lewis' ante rem modal realism), its distinguishing feature being that the existence of worlds is contingent upon the existence and activity of rational agents:

> [T]he moderate realist believes that the only possible
> worlds there are—ways things might have been—

[1]From the preface to *On the Plurality of Worlds*, modal realism is "the thesis that the world we are part of is but one of a plurality of worlds, and that we who inhabit this world are only a few out of all the inhabitants of all the worlds." (Lewis 1986, p. vii)

[2]Lewis (1986) and Van Inwagen (1986) agree that such representationalism comes in three varieties: linguistic, pictorial, and magical. Van Inwagen defends the last variety (but calls it 'unsound' rather than 'magical').

[3]"The other worlds are of a kind with this world of ours. To be sure, there are differences of kind between things that are parts of different worlds... but these differences of kind are no more than sometimes arise between things that are parts of one single world... The difference between this and the other worlds is not a categorial difference." (Lewis 1986, p. 2).

> are (like everything that exists at all) elements of our actual world. They obviously are not concrete objects or situations, but abstract objects whose existence is inferred or abstracted from the activities of rational agents. (Stalnaker 1984, p. 32)

Clearly, ME lies somewhere in the relevant position space between MR and a strong actualism that posits no plurality of worlds, whether concrete and necessary or abstract and contingent. But there is an illuminating sense in which MR and Stalnaker's ME are cousins within a theoretical lineage, namely, in their reductive treatment of actuality, i.e., the "indexicality theory" of actuality (IT). According to IT, what it means for a world to be actual is just like what it means for a time to be now: Every time is now relative to itself and not to any other time; likewise, every world is actual relative to itself but not to any other world. Stalnaker argues that "the semantic thesis that the indexical analysis of 'actual' is correct can be separated from [MR]" (Stalnaker 1984, p. 29), a claim that, if true, clearly pushes ME towards MR on the position spectrum from radical possibilism to radical actualism.

However, there are other ME-ists who do not subscribe to IT. Robert Adams, for example, rejects IT on the grounds that it is implausible that ascribing actuality to an event (e.g., by saying "this happened") involves assigning it a location in modal space, and instead proposes that "actuality is a simple, unanalyzable property of the actual world, by which it is distinguished from the other possible worlds" (Adams 1974, p. 221). This "simple property theory of actuality" pushes ME in the direction of actualism, and Adams, indeed, explicitly refers to his view as "actualism about actuality."[1] Similarly, Peter van Inwagen, presumptively ruling out MR, claims that "no one who accepts an 'abstractionist' [i.e., ersatzist] modal ontology can accept any account of actuality that could possibly be described as 'indexical'" (van Inwagen 2001, p. 10). In place of IT, van Inwagen proposes that a world is "a possible proposition that, for any proposition p, entails either p or the denial of p;

[1]Nevertheless, Adams further qualifies his "actualism" as committed to the existence of "furniture [within the actual world] rich enough for the logical construction of a plurality of completely determinate possible worlds" (Adams 1974, p. 225). Adams opposes his "soft actualism" (SA) to "hard actualism" (HA) according to which talk of possible worlds is at most "a heuristic device in thinking about theories and problems in modality" (Adams 1974, p. 224).

and actuality is simply truth" (van Inwagen 1980, p. 230); hence a possible world on van Inwagen's view is a maximal proposition, and the actual world is the (one and only) true maximal proposition.

So there are two general camps concerning the nature of actuality: Lewis-Stalnaker and Adams-van Inwagen. The first camp reduces actuality to an index: being actual is like being now; the man "who says that his world alone is the actual world is as foolish as a man who boasts that he has the special fortune to be alive at a unique moment in history: the present" (Lewis 1970, p. 20). The second camp holds that actuality is irreducible—an unanalyzable property that a world can have and that one world does in fact have, or else it's just plain old truth. Within the first camp are MR-ists like Lewis and ME-ists like Stalnaker. Within the second camp are ME-ists like van Inwagen and Adams as well as philosophers who reject possible-worlds talk out of hand. For the purposes of this paper, the first camp is possibilism and the second is actualism. Admittedly, this terminology goes against the convention that "possibilism" is synonymous with "modal realism" and actualism encompasses every other theory (see fn. 1); but it is a motivated heterodoxy. In order to avoid circularity, the actualist account of possible worlds as sets of propositions must not reduce propositions to sets of possible worlds — a point van Inwagen appreciates in declaring that a proposition "is not...a set-theoretical construct on sets of extended objects" (van Inwagen 1980, p.230), but rather an irreducible intensional entity.

## 2. Possibilist Reduction of Intensional Entities

We have seen that a reductionist theory, IT, is central to possibilism. A second reductionist theory, a reduction of properties and propositions to set-theoretic constructs of possibilia — call it 'ST'— is central to possibilism as well; moreover, as we will see in §§3-4, it is ST that is the source of the internal tension within possibilism alluded to in the introductory remarks. Propositions, on ST, are sets of possible worlds: the proposition that p is the set of worlds at which p (or the function that maps the p-worlds to truth and the ~p-worlds to falsehood). Properties are sets of possible individuals; the property F is the set of all x such that Fx (or the function that maps all the worlds to all the Fs within them). Part of the motivation behind ST is a taste for extensionalism, i.e., for entities like sets that are identical if consisting in the same members (hence, identical if coextensional), rather than intensional entities whose identity conditions are difficult to specify (co-extension is likely necessary but insufficient). But ST has a further motivation: Actualists who countenance the existence of possible worlds as abstract objects have a reductive theory of their own, a reduction of possible worlds to propositions, and possibilists have principled objections to the actualists' reductionism. As we will see, ST and the corresponding actualist theory are something like mirror-images, with more or less the same items flanking an equals-by-analysis sign, but with the roles of analysandum and analysans reversed.

Adams gives an archetypal statement of the actualist reduction of possible worlds: "Let us say that a world-story is a maximal consistent set of propositions. That is, it is a set which has as its members one member of every pair of mutually contradictory propositions" (Adams 1984, p. 225). The process commences, with each even-numbered step taken as a reductive analysis of the preceding odd-numbered step:

> (1) There is a possible world in which p.
> (2) The proposition [p][1] is a member of some world-story.
> (3) In every possible world, q.
> (4) The proposition [q] is a member of every world-story.
> (5) Let w be a possible world in which r. In w, t.
> (6) Let s be a world-story of which the proposition [r] is a member. The proposition [t] is a member of s. (ibid.)

One can easily see how the need for talk of worlds falls away entirely. For (2), there is (2a): The proposition [p] is a member of a maximal consistent set of propositions. For (4), there is (4a): The proposition [q] is a member of every maximal consistent set of propositions; from which follows (4b): the proposition [~q] is a member of no maximal consistent set of propositions; so q is something

[1]Substituting our proposition notation for Adams'

like 2+2 = 4, and ~q is something like 2+2 ≠ 4. The idea generalizes; by means of this reduction, actualists can avoid talking about possible worlds altogether, and do all the work possible worlds are supposed to do with traditional modal operators and facts or propositions.

Stalnaker gives as a reason for reversing the order of the analysis the fact that "[w]hat is more important than familiarity[1] for choosing the order of an analysis is the question of structure":

> Normally, one would expect to analyze the more highly structured set of entities in terms of a less highly structured one so that the analysis might help explain the structure. Propositions are highly structured... Possible worlds, on the other hand, are relatively unstructured. (Stalnaker 1984, p. 35).

I think the point is a valid one, though it tells against Stalnaker's ME and in favor of MR. If we should order analyses such that more highly structured entities are the analysanda and less highly structured ones are the analysantes, then surely, in ideal circumstances, the terminal nodes of an analysis would be without logical structure. And while ME-ist possible worlds are "relatively unstructured," MR possible worlds, being concrete particulars, are surely without logical structure absolutely. A further advantage of reducing propositions to worlds, again from Stalnaker, is that doing so provides a closure principle: "For every set of propositions Γ there is a proposition A such that Γ implies A, and A implies every member of Γ" (ibid.). Whereas the world-story reduction leaves the maximal consistency condition as a primitive, and perhaps a problematic one.

If there are, prima facie in any case, some advantages in going the possibilist route in analysis, from propositions to set-theoretic constructs of worlds, are

[1]The "familiarity" referred to is our supposed greater familiarity with propositions and facts than with possible worlds. But considerations of familiarity only tell against ST given ME; a clear advantage of MR relative to ME, in this case, is that *concrete* possible worlds are utterly ordinary sorts of things: just big rocks, and if you've seen one, you should be able to recognize another. However no one ever bumps into a fact, nor, unlike with entities like electrons, can there even be experiments to detect them. So for MR-ists, considerations of familiarity as well as structure tell in favor of the possible worlds reduction.

there drawbacks to doing so as well? In fact, there are more credible objections to the possible worlds account of propositions than the scope of this paper permits investigating. I will consider what I think are the two strongest objections to the possible worlds account. First, the possible worlds account entails not only a closure principle, which may be construed as a fairly unqualified good, but a simple identity condition for propositions: Equivalent propositions are identical, since propositions are sets (or set-theoretic constructs); sets are extensional entities, and equivalent extensional entities are identical. The practical consequences of this entailment pose grave problems for the theory; all the necessary truths (including, uncontroversially, the mathematical truths and the logical truths), for example, turn out to be identical. But surely not! The proposition that 2+2 = 4 is surely not the same proposition as the proposition that if x exists, x is self-identical. Yet both hold at all possible worlds, and therefore if they are the sets of worlds at which they hold, they are one and the same set. (The problem recurs if propositions are world-to-truth-value functions.)

The possibilist can get out of this dilemma by refining the analysis to hold that a proposition is not a set of worlds (or a function from worlds to truth-values, which would face the same difficulty vis-a-vis necessary propositions), but rather an ordered sequence. In other words, the possible worlds account morphs into a propositional-complex reduction of proposition. This move faces trouble of its own:

> For example, the proposition that you dream is the ordered set <dreaming, you>...Similarly, the proposition that you dream and I think is the ordered set <conjunction, <dreaming, you>, <thinking, me>>; the proposition that someone dreams is the ordered set <existential generalization, dreaming>; and so forth... The...concern with the propositional-complex theory is that there is no way to determine which ordered set is the alleged item I believe. Is it <dreaming, you>? Or is it <you, dreaming>? The choice is utterly arbitrary. Admitting this kind of wholesale arbitrariness into a theory would be

> unwarranted if there were an otherwise acceptable alternative which is free of it. (Bealer 1998, pp. 6-7)

In other words, the avenues of escape for the possibilist appear to be blockaded. To avoid the result that all necessary propositions are identical, the possibilist must resort to a Chisolming of her theory that entails objectionable arbitrariness. And if arbitrariness of this sort is deemed fatal to the theory that entails it, and if there is no means of rendering the arbitrariness benign through further Chisolming (say, via epistemicism), then the possible worlds reduction is doomed[1]—if not, anything resembling the traditional account of propositions as the objects of belief, desire, etc. might have to be abandoned. Note, though, that George Bealer places a condition on jettisoning the propositional-complex reduction, namely, if there were an otherwise acceptable alternative.

Is there an otherwise acceptable alternative? Bealer's is to take propositions (and properties) as irreducible, sui generis intensional entities. If we are prepared to countenance the existence of such things, then Bealer's alternative is in excellent position to carry the day. But one of the virtues, for possibilists, of the possible-worlds reduction, is precisely that it promises an escape from ante rem realism about intensional entities; in Quine's (1968) memorable phrasing, the possible-worlds

[1]No less an expert witness than David Lewis himself testifies to the fatal objectionability of an indeterminacy that would appear to be of a kind with that created by any sort of propositional-complex reduction (whether it further reduces the term in any argument space or not). Regarding the theoretical possibility of restricted (versus unrestricted) mereological composition, Lewis writes: "There is such a thing as the sum, or there isn't. It cannot be said that, because the *desiderata* for composition are satisfied to a borderline degree, there sort of is and sort of isn't. What is this thing such that it sort of is so, and sort of isn't, that there is any such thing?... Restrict quantifiers, not composition. *Vague existence, speaking unrestrictedly, is unintelligible; vague existence, speaking restrictedly, is unproblematic*" (Lewis 1986, pp. 212-213) [emphasis mine]. The distinction Lewis raises is between vagueness as a linguistic phenomenon and vagueness as an ontological phenomenon. The former is not problematic, indeed ordinary language is rife with it; whereas as the latter is unintelligible. Bealer would agree: "This ontological question [of what propositions are] differs from the model-theoretic question of whether extensional entities might be used merely to *represent* propositions" (Bealer 1998, p. 4). In other words, it is unproblematic for a model to entail an arbitrary choice between representing the proposition that I dream as <dreaming, I> or <I, dreaming> but it cannot be the case that the proposition, *i.e.*, that thing, is indeterminate between the two. There is, however, a third option that neither Lewis nor Bealer mentions: epistemicism, the view that in such cases, unbeknownst to us, there is a fact of the matter as to which ordered set the proposition *is*. Epistemicism, to be sure, raises a host of new difficulties, which I am in no position to examine here. But it is an option. For the possibilist, it might be the only alternative to abandoning propositions as objects of belief, desire, knowledge, etc.

reduction "stays within a clear extensional ontology." One might note that Bealer's (and Lewis') objection to ontological indeterminacy (see fn. 9) is at least superficially similar to Quine's (and Lewis') objection to intensional ontology; they are both objections by appeal to metaphysical intuitions. Is ontological indeterminacy more or less gruesome, on naive metaphysical intuition, than ante rem realism about intensional entities?[1] I am inclined to say ontological indeterminacy is worse, but not quite decisively so; moreover, the epistemicist escape hatch for the ordered set reduction, whatever one thinks of epistemicism, surely adds something to the possibilist side of the ledger.

In any case, both iterations of the possible worlds reduction face a second objection, which further applies, mutatis mutandis, to the possibilist reduction of properties to set-theoretic constructs of possible individuals:

> For example, the proposition that I dream = the function that maps possible worlds in which I dream to the true and other possible worlds to the false. And a sensible property such as the aroma of coffee = the function that maps possible worlds to the set of things which have the aroma of coffee in that world... When I believe (doubt, justify, assert) some proposition, do I believe (doubt, justify, assert) a function? On the face of it, this is not plausible. Advocates of this reduction seem to have lost "the naive eye." (Bealer 1998, pp. 4-5)

Needless to say, the prima facie plausibility complaint applies just as naturally to the formulation according to which propositions and properties are sets, ordered or otherwise, rather than functions. The point is that the possibilist is missing the foundational motivation of the theory of propositions, namely, that they are the

[1]Bealer (1998) has an argument for the necessary existence of properties from intensional logic, specifically, from consideration of "transmodal propositions." I am inclined to accept this argument as decisive for actualists, though not necessarily for possibilists (which is the whole game for present purposes). Additionally, Moffett (2002) poses challenges to the theory of 'that'-clauses as singular terms upon which the argument from intensional logic depends.

objects of those attitudes, like believing, desiring, knowing, doubting, etc., that are eponymously termed ‘propositional attitudes’. On the traditional theory, all such attitudes are simple two-place relations holding between a subject who is some kind of conscious agent and an object, which is a proposition, and propositions are just that, propositions, not sets (ordered or otherwise or functions). One believes that-something, one does not believe the set such that thus and such. Or so the story goes; I can imagine the Quinean response to the effect that, of course extensional entities are the objects of belief — since there are objects of belief, and the only objects there are, are extensional.[1]

Thus the possibilist extensional reduction is complete. Let us remain, for the time being, agnostic about whether or not possibilism can survive the refutations adduced here in §2. What the foregoing criticisms had in common is that they were marshaled on behalf of actualism and pointed toward actualist conclusions. What I wish to concentrate on is a problem that possibilism creates for itself, without any intrusion of actualist assumptions or premises. To wit: possibilism’s extensional reduction is, prima facie, inconsistent with another of its theoretical commitments.

## 3. Possibilism and Fregean Semantics

Possibilism is broadly Fregean; in other words, it adopts the broad features of the theory of linguistic meaning first enunciated in Gottlob Frege’s “Über Sinn und Bedeutung” (translated alternately as “On Sense and Reference” and “On Sense and Nominatum”), according to which there are two constituents of the meaning of an expression. In Frege’s initial formulation:

> If the sign ‘a’ differs from the sign ‘b’ only as an

[1]Likewise, possibilists can escape essential-indexical type worries by means of Lewis’ division of attitudes into *de dicto* and *de se*, whereby “[b]elief *de dicto* is self-locating belief with respect to logical space; belief irreducibly *de se* is self-locating belief at least partly with respect to ordinary time and space, or with respect to the population” (Lewis 1979, p. 140). The upshot is that beliefs (since on this account *de dicto* belief is reducible to *de se* belief) turn out not to be standard two-place relations between believers and propositions but self-predications of properties. Suspending judgment on the plausibility of this reduction, the relevant point for our purposes is that properties, too, are subject to an extensional reduction.

> object (here by its shape) but not by its role as a sign, that is to say, not in the manner in which it designates anything, then the cognitive significance of "a = a"would be essentially the same as that of "a = b", if "a = b" is true. A difference could arise only if the difference of the signs corresponds to a difference in the way in which the designated objects are given. (Frege 1892 in Martinich, p. 199)

Take any pair of co-denoting names; say, 'George Orwell' and 'Eric Blair.' Both of the sentences, 'Orwell = Orwell' and 'Orwell = Blair' are true, as are the propositions they express, [Orwell = Orwell] and [Orwell = Blair]. But the first sentence patently does not have the same meaning as the second sentence, since the first sentence expresses the first proposition, the second sentence expresses the second proposition, and the two propositions are clearly not the same. Yet 'Orwell' and 'Blair' pick out the same individual, namely, him. Therefore, Frege argues, there must be more to the meaning of a name than its reference, and this second constituent of meaning he termed 'sense': "A proper name (word, sign, sign-compound, expression) expresses its sense, and designates or signifies its nominatum [reference]" (ibid., p. 202).

Fregeanism has in its favor a further virtue that direct reference theory does no-t—it can give a powerfully simple and persuasive account of the phenomenon of intensionality in language, without having to resort to pragmatics. To wit: Take Jones, who is acquainted with the principle of self-identity but knows nothing of the biography of Orwell/Blair. When asked if Orwell is Orwell, Jones answers 'yes'; when asked if Orwell is Blair, Jones answers 'no.' Hence 'Jones believes Orwell is Orwell' is true while 'Jones believes Orwell is Blair' is false. Yet if direct reference theory is correct, 'Jones believes Orwell is Orwell' and 'Jones believes Orwell is Blair' have the same meaning, and so it cannot be the case that the two sentences differ in truth value; so provided Jones truthfully answers the first question affirmatively, he speaks falsely when he answers the second question negatively. And this conclusion strains credulity; Jones knows nothing about Orwell or Blair, let alone that Orwell = Blair, yet by the lights of direct reference theory, simply in virtue of understanding the principle of self-identity, Jones does know that Orwell = Blair, his protestations of complete ignorance on the question notwithstanding. Perhaps Jones was the logician who discovered the principle of self-identity, and needless to say, lived and died long before Orwell/Blair was born. So Jones knows as well as anyone else who has ever lived that if x exists, x = x. So when Jones is informed that Blair (tense-neutrally) exists, Jones undoubtedly knows that Blair = Blair. And therefore, according to direct reference theory, Jones knows that Orwell = Blair before Orwell was born, learned English, and decided to adopt the pseudonym 'Orwell'. Is this plausible?

Belief reports, moreover, are only one instance of the failures of salva veritate substitution that seem prima facie to embarrass direct reference theory. David Chalmers nicely encapsulates a related embarrassment of direct reference theory:

> The sentence 'Hesperus is Hesperus' is trivial. It can be known a priori, or without any appeal to experience. The knowledge that Hesperus is Hesperus requires almost no cognitive work at all, and gives us no significant information about the world. By contrast, the sentence 'Hesperus is Phosphorus' is nontrivial. It can only be known a posteriori, by appeal to empirical evidence. The knowledge that Hesperus is Phosphorus requires much cognitive work, and gives us significant information about the world. (Chalmers 2002, p. 137)

To be sure, direct reference theorists have a supply of strategies for explaining away such massively counterintuitive results.[1] Let us leave them to that task,

[1]Unfortunately, their strategies are riddled with flaws that further weaken and complicate direct reference theory (see Bealer 2004). And perhaps most devastating of all, even if any effort at explaining away substitution failure can survive its refutation, direct reference theory "has the absurd consequence that the answer to *any* question that one might entertain can be known right now by means of a few easy steps (for example, the question whether there are extraterrestrials)" (Bealer, 2005) — at least, it has this consequence if the direct reference theorist insists on construing knowledge as a two-place relation. It might be avoided, at the cost of adopting a counterintuitive and problematic view of knowledge as a three-place relation.

noting for our purposes that there is a compelling argument for Fregean semantics and against direct reference theory that is utterly orthogonal to possibilism versus actualism.

Yet possibilism very definitely belongs to the Fregean family, whereas actualists can be found in large numbers on both sides of the debate over Fregean semantics. I want to make a stronger claim: It is no happenstance that possibilism is aligned with Fregean semantics; direct reference theory is to varying degrees incompatible with possibilism, such that the more thorough-going possibilism is in its extensionalism, the greater its incompatibility with direct reference theory. This incompatibility comes to the fore when possibilists of either the MR or the ME stripe attempt to cash out the possibilist understanding of the tenets of modal logic. The most thoroughly extensionalist such effort is Lewis' counterpart theory, the first six postulates of which entail a peremptory rejection of direct reference theory[1]:

> P1: $\forall x \forall y(Ixy \supset Wy)$
> (Nothing is in anything except a world.)
> P2: $\forall x \forall y \forall z(Ixy \,\&\, Ixz \supset y=z)$
> (Nothing is in two worlds)
> P3: $\forall x \forall y(Cxy \supset \exists z Ixz)$
> (Whatever is a counterpart is in a world)
> P4: $\forall x \forall y(Cxy \supset \exists z Iyz)$
> (Whatever has a counterpart is in a world)
> P5: $\forall x \forall y \forall z(Ixy \,\&\, Izy \,\&\, Cxz \supset x=z)$
> (Nothing is a counterpart of anything else in its world)
> P6: $\forall x \forall y(Ixy \supset Cxx)$
> (Anything in a world is a counterpart of itself)
> (Lewis 1968, p. 27)

There are two more postulates which do not bear on the conflict between possibilism and direct reference theory, and one more postulate that Lewis might have included given the later development of his views, viz., a postulate to rule out

[1]In Lewis' notation, 'Wx' means x is a possible world, 'Ixy' means x is in possible world y, and 'Cxy' means x is a counterpart of y.

impossible objects. Call it 'P9': $\forall x(\exists y(Ixy))$—for any x, there exists a y such that x is in y. From P1 and P9, it follows that everything is in a world. (P1 alone is not sufficient to rule out impossible objects, since it requires only that something that is in something be in a world and nothing else, but does not rule out that something could exist that is not in anything, even a world, and hence is an existing impossible object.) For the sake of insurance, one could also posit P10: $\forall x(Wx \supset Ixx)$ — for any x, if x is a world, x is in itself; i.e., worlds are improper parts of themselves. The idea is to pre-empt a mistaken interpretation of P9 to the effect that P9 rules out the existence of the worlds, which would have the unhappy consequence of forcing the counterpart theorist to choose between affirming the existence of impossibilia or else simply gainsaying possibilism; P10 blocks that dilemma.

From the conjunction of P1-P6 and P9-P10, it follows that everything is either a world or a proper part of a world, and everything is world-bound, existing exclusively at one world. These consequences of counterpart theory render it fundamentally incompatible with direct reference theory, since on direct reference theory, for any world w, 'Humphrey' denotes Humphrey or nothing. But on counterpart theory, Humphrey exists only in the actual world. So the intuitive modal statement, 'Humphrey might have won' turns out not to be true given the postulates of counterpart theory and the premises that names refer directly and are rigid designators. What must be the case, given counterpart theory, for 'Humphrey might have won' to be true, is that the name 'Humphrey', when used in a modal context, denotes not Humphrey, but an individual who stands in the counterpart relation to Humphrey, transmodal instances of that relation being incompatible with the identity relation.[1] Hence, some kind of qualitative similarity relation short of identity must fix the reference of names in modal contexts. And, on grounds of parsimony, the counterpart theorist is driven to the conclusion that reference in general is fixed by qualitative features.[2]

So it should be safe to say, qualifying the statement with which I began this section, that possibilists who adopt counterpart theory are committed to some broadly Fregean semantic theory. Which raises the question: Are possibilists

[1] This can be derived from P2 and P5: Given that $\forall x\forall y\forall z(Ixy \,\&\, Ixz \supset y=z)$ and $\forall x\forall y\forall z(Ixy \,\&\, Izy \,\&\, Cxz \supset x=z)$, it follows that $\forall w\forall x\forall y\forall z(Cwx \,\&\, Ixy \,\&\, Iwz \,\&\, y\neq z \supset w\neq x)$.

[2] The conclusion can be further supported by examples that cry out for associating a name with a descriptive sense, e.g., color names (see Lewis 1997, p. 353n and *inter alia*).

committed to counterpart theory? The answer is, while it is technically coherent to construct a possibilism that is not counterpart-theoretic, the resultant theory will turn out to clash with whatever intuitive support there is for the reductive tenets of possibilism described in §§1-2. The alternative to counterpart theory relies on the view known as "haecceitism,"[1] according to which, in addition to whatever qualitative properties an object has, it has a non-qualitative property, a brute thisness or haecceity, such that it can be individuated from a qualitative indiscernible. (Haecceitism thus provides grounds for denying the identity of indiscernibles.) If a possibilist were to adopt haecceitism, she could say that Humphrey does indeed exist at multiple possible worlds, that the term 'Humphrey' in 'Humphrey might have won' denotes Humphrey at whatever worlds in which he did win, and that what fixes the reference of Humphrey is his haecceity, not any qualitative or descriptive feature; in other words, direct reference via primitive (or at least non-qualitative) transworld identity.

This is a coherent position that a possibilist could take. But it is an unmotivated position: Haecceitism divides the properties into two classes, the non-qualitative properties, namely the haecceities, and all the others. Yet we are still operating from within a possibilist framework; so all properties, qualitative or not, are subject to a set-theoretic reduction; so, like any other property, the haecceities are reduced to a set-theoretic construct. What results is a division of extensional set-theoretic constructs into a class of qualitative set-theoretic constructs and a class of non-qualitative constructs. This distinction is not just primitive, it is utterly mysterious. What could possibly account for it being the case that some sets (ordered sets, functions, whatever the formulation is) constitute qualitative properties and some sets constitute non-qualitative properties? Within a metaphysic of extensional reduction, such a distinction is gruesome. The only rationale for a possibilist to reject counterpart theory and adopt haecceitism is a desire to reconcile possibilism with direct reference theory. But that is to beg the question. And we have already seen at some length that direct reference theory is beset by problems. So the possibilist should be a counterpart theorist; so the possibilist should be a Fregean. Consequently, possibilists have no alternative but to endeavor to solve Frege's

[1]See Kaplan 1975 and Lewis 1986, §4.4 and *inter alia*.

puzzle head on,[1] and if their theories are to be successful, they must discover the senses whose elusiveness is a prime motivation for direct reference theory.

## 4. A Possibilist Solution

Possibilism subjects ontology to an extensional reduction: the intensional entities are reduced to set-theoretic constructs of possible worlds and possible individuals. In light of that extensional reduction, there are no transworld referring terms available to make the intuitive modal statements come out true; so recourse to a counterpart theory that depends upon some kind of Fregean semantics is unavoidable. But the commitment to a Fregean semantics further entails that for possibilists, the classic substitution failures underlying Frege's puzzle and similar puzzles are to be explained by intensionality genuinely inhering in linguistic meaning. In other words, for the possibilist, 'a=a' and 'a=b' literally and unambiguously differ in meaning, and the failure of substitution salva vertitate of 'b' for 'a' is not the result of some kind of communicative or pragmatic opacity, but of the meanings of the names being unidentical though coextensional. The prima facie inconsistency within possibilism should now be apparent: the paradigmatic candidates for senses, properties and the like, have been reduced to extensional entities, i.e., entities that are identical if coextensional. If 'a=a' and 'a=b' differ in meaning because the property of being a differs from the property of being b, but the property of being a is the set of the a's, and the property of being b is the set of the b's, then we arrive at a contradiction, as 'a=a' and 'a=b' turn out to have the same meaning. Or consider the puzzle propositionally: If [a=a] is the set of worlds at which a=a, and [a=b] is the set of worlds at which a=b, then it would seem that the propositions expressed by 'a=a' and 'a=b' are identical. The question before us, then, is whether we can account for substitution failures within a Fregean framework, without repudiating the reduction of propositions and properties to set-theoretic constructs.

A thesis, counterintuitive at first glance, that counterpart theory suitably worked out would nevertheless seem to entail, sheds some light on a possibilist strategy for

[1]Direct reference theory simply evades the puzzle, by claiming that, even when A=B, 'A=A' and 'A=B' do not differ in meaning, and the substitution failures Frege and his successors adduce are to be explained away by pragmatics. This evasion is the source of the difficulties noted above (see fn. 6).

doing the work with extensional entities that intensional entities have traditionally done. That thesis is the contingency of identity:

> I say that the plastic and the dishpan are identical. I also say they might not have been...And what I mean is not the absurd falsehood that something, namely the plastic which is the dishpan, has a de re possibility of failing to be self-identical. What I mean is that an identity pair has the de re possibility of being a non-identity pair. (Lewis 1986, p. 263)

Contingency of identity roughly amounts to a denial of a principle implicit in my reconstruction of the propositional version of Frege's puzzle: $(\forall xy)\ x{=}y \Rightarrow \Box x{=}y$. This principle can be derived from Leibniz's Law, $x{=}y \Rightarrow (\forall F)(Fx \Leftrightarrow Fy)$ and the seemingly innocuous premise that any y has the property of being necessarily identical to y. It follows then that if x=y, since (by Leibniz's Law) x has every property y has, x has the property of being necessarily identical to y. The way to block this derivation, and hence the necessity of identity, is to note that on counterpart theory, which rules out the existence of individuals across modal space, being necessarily identical to is not a genuine property of individuals, but a combination (for lack of a better term) of a property (identity) with an operator (necessity). Thus, the sentence we would render in English as "x is necessarily self identical" is an operator sentence, $\Box[x{=}x]$, not a predicative sentence $N[x{=}x]$.[1] The operator sentence is true: at all worlds, everything is self-identical; but the predicative sentence is not true: nothing has the property of being necessarily self-identical, because there is no such property of individuals.

What makes contingent identity relevant to substitution failure is that the possibility of the contingency of identity—i.e., the existence of an identity pair

[1]Bealer explicitly reverses this formula: "[O]ur considerations about logical form showed that [It is necessary that A] is *not* an operator sentence...but is instead a *predicative* sentence" (Bealer 1998, p. 138). The counterpart theorist *must* argue, whatever the similarity in form between "It is necessary that A" and genuine predicative sentences, that sentences ascribing necessity to singular terms (that-clauses, names of individuals, etc.) can never be predicative. This position is not mere gainsaying; it rises or falls with the denial of transworld existence.

<x,y> that stands in the counterpart relation to a non-identity pair <u,v>—by no means entails that all identities are contingent. (And to be sure, the identities that are contingent are necessarily contingent, the identities that are necessary are necessarily necessary.) Hence, the intensionality that is preserved in a fully extensional ontology may be interpreted as consisting, in the case of individuals, in reference to a non-identity counterpart pair of an identity pair, and in general cases, in a distinction between necessary and contingent co-extension. That is, given an ontology constituted entirely by individuals and set-theoretic constructs of individuals, of the set-theoretic constructs that are coextensional at any given world, there are some constructs that can be shown to be non-coextensional at other worlds. Take, for example, the properties of being a creature with a kidney (a renate) and a creature with a heart (a cordate): At the actual world, the proposition that renates are cordates is true; in other words, the actual world is a member of the set of worlds at which every renate is a cordate and every cordate is a renate. But at many other possible worlds, there are renates that have no heart and cordates that have no kidney. At these worlds, the proposition that renates are cordates is false. The contingency of the coextension of the property of being a renate and the property of being a cordate at any worlds at which they are in fact coextensional supplies the possibilist with an account of substitution failures between the two. When we predicate across all of modal space without restriction, the proposition that renates are cordates is simply false, because the set of all creatures at all worlds with kidneys and the set of all creatures at all worlds with hearts have different membership. When we restrict predication to our world, it turns out as an empirical matter that the set of renates and the set of cordates have the same membership. But because their coextensionality is contingent, they cannot be substituted in all contexts salva veritate.

On the other hand, consider the property of being a creature with a kidney and the property of being a renate, which I used interchangeably above. There are, presumably, no worlds at which any individual is a renate but not a creature with a kindey, nor a creature with a kidney but not a renate. So the coextension of the properties is necessary: at any worlds at which there are creatures with kidneys, those individuals are renates, and at any worlds at which there are renates, those individuals are creatures with kidneys. In cases of necessary coextensions, we can conclude that the coextensional set-theoretic constructs serving as Fregean senses are in fact identical. So the property of being a creature with a kidney is the property of being a renate. And this should accord with our intuition that in the common cases of substitution failure we have been addressing, there will be no substitution failure for these properties. E.g., if c believes that b is a creature with a kidney, then c believes that b is a renate; and if c believes that b is a renate, then c believes that b is a creature with a kidney. Belief reports regarding the ascription of these properties are not subject to substitution failure, and indeed, were c to report himself as believing that b is a renate but not a creature with a kidney, we would likely conclude that he is simply misunderstanding terms.

So we are in a position to state a theory of intensionality within a fully extensionally reduced modal space: intensionality consists in set-theoretic constructs standing in a relation of contingent coextensionality to some set-theoretic constructs and necessary coextensionality to others; context will determine whether any set-theoretic constructs are necessarily or contingently coextensional, and therefore, will determine when substitution failure occurs and when it does not. Note that this view does not involve an illegitimate predication as the case of necessary identity above did, because in the preceding case, necessary identity was ascribed to individuals, rather than set-theoretic constructs of individuals. No individual (on counterpart theory) exists in more than one world; but transworld sets exist, simply because sets exist in virtue of their members existing, so there is a set whose members are any two or more individuals who exist at different possible worlds. Set-theoretic constructs can stand in relations of necessary and contingent coextensionality while individuals cannot have the property of being necessarily self-identical — but it is nevertheless true at all possible worlds that everything is self-identical.

The theory proposed, therefore, is in effect an extensional reduction of intensionality; intensionality is to be understood as coextensionality between any two set-theoretic constructs over a range smaller than all possibilia. Is this view a plausible account of what intensionality in fact is? Its merits and demerits are those of possibilism, and it plays by the possibilist rule that nothing is to be assigned as the value of a bound variable, as it were, but individuals and set-theoretic constructs thereof. The theory of intensionality as coextensionality over a restricted range serves, at least, to supply possibilism with an account of substitution failure, and does so without introducing any new problematic complications to possibilism. If it is to be rejected on plausibility grounds, then so, in all likelihood, should possibilism be rejected.

## Works Cited

Adams, Robert Merrihew. "Theories of Actuality." *Nous* 8 (1974): pp. 211-31.

Bealer, George. "Universals." *Journal of Philosophy* 90 (1993): pp. 5-32.

———. "An Inconsistency in Direct Reference Theory." *Journal of Philosophy* 111 (2004): pp. 574-93.

———. "A Proof Within Direct Reference Theory That You Are Omniscient." Unpublished paper, 2005.

———. "Propositions." *Mind* 107 (1998): pp. 1-32.

———. "Universals and Properties." In *Contemporary Readings in the Foundations of Metaphysics*, Laurence, Stephen and Macdonald, Cynthia, eds. pp. 131-147. Oxford: Basil Blackwell, 1998.

Chalmers, David. "On Sense and Intension." *Philosophical Perspectives* 16 (2002).

Frege, Gottlob. "On Sense and Nominatum" [1892]. In *The Philosophy of Language*, edited by A.P. Martinich, pp. 199-211. New York: Oxford University Press, 2001

Lewis, David. "Counterpart Theory and Quantified Modal Logic" [1968]. Now reprinted in Lewis, *Philosophical Papers*, vol. 1. New York: Oxford University Press, 1983, pp. 26-46.

———. "Anselm and Actuality" [1970]. Now reprinted in Lewis, *Philosophical Papers*, vol. 1. New York: Oxford University Press, 1983, pp. 10-20.

———. “’Tensions” [1974]. Now reprinted in Lewis, *Philosophical Papers*, vol. 1. New York: Oxford University Press, 1983, pp. 250-260.

———. “Attitudes De Dicto and De Se” [1979]. Now reprinted in Lewis, P*hilosophical Papers*, vol. 1. New York: Oxford University Press, 1983, pp. 133-160.

———. *On the Plurality of Worlds*. Oxford: Basil Blackwell, 1986.

———.“Naming the Colours” [1997]. Now reprinted in Lewis, *Papers in Metaphysics and Epistemology*. New York: Cambridge University Press, 1999, pp. 332-358.

Quine, W.V. “Propositional Objects” [1968]. In *Ontological Relativity & Other Essays*. New York: Columbia University Press, 1969, pp. 139-160.

Stalnaker, Robert C. “Possible Worlds” [1984]. Now reprinted in Stalnaker, *Ways a World Might Be*. Oxford: Oxford University Press, 2003, pp. 25-39.

———. “Counterparts and Identity” [1987]. Now reprinted in Stalnaker, *Ways a World Might Be*. Oxford: Oxford University Press, 2003, pp. 111-132.

———. “Comparing Qualia Across Persons”[2000]. Now reprinted in Stalnaker, *Ways a World Might Be*. Oxford: Oxford University Press, 2003, pp. 219-238.

———. "On Considering a Possible World as Actual" [2001]. Now reprinted in Stalnaker, *Ways a World Might Be*. Oxford: Oxford University Press, 2003, pp. 188-200.

Van Inwagen, Peter. "Indexicality and Actuality" [1980]. Now reprinted in van Inwagen, *Ontology, Identity, and Modality*. Cambridge, U.K.: Cambridge University Press, 2001, pp. 165-185.

———. "Two Concepts of Possible Worlds" [1986]. Now reprinted in van Inwagen, *Ontology, Identity, and Modality*. Cambridge, U.K.: Cambridge University Press, 2001, pp. 206-242.

# The Cognitive Command Constraint in Wright's "Truth and Objectivity"

Peter Goldstein
Cornell University

In "Truth and Objectivity," Crispin Wright develops a framework for carrying out debates about realism and anti-realism. His framework employs the idea that whether or not we should think realistically about a discourse depends on how the notion of *truth* operates in the discourse. Wright himself endorses a *pluralism* about truth, arguing that any predicate in a given discourse that satisfies a set of minimal platitudes *thereby* qualifies as a truth predicate for that discourse. The question about whether we should be realists about a given minimally truth-apt discourse then amounts to the question of whether the (minimal) truth predicate of that discourse satisfies further, realism-relevant, features. One such feature is *Cognitive Command*: we say that a discourse exhibits Cognitive Command just in case it is a priori that disputes occurring within the discourse must (modulo certain provisos) be attributable to something worth regarding as a cognitive shortcoming. This constraint on the minimal truth predicate operative within the discourse is motivated by the idea that if a discourse exhibits Cognitive Command, then it is natural to think of true assertions within that discourse as "statements of fact" that are true or false depending on how well they "represent the objective states of affairs" that are the subject matter of that discourse.

In this paper, I will discuss Cognitive Command, and examine various arguments purporting to show that the condition is satisfied by any discourse sustaining the definition of a minimal truth predicate. This paper is divided into four sections. In section **(1)**, I will discuss Wright's *minimalism*. First I will discuss the set of minimal platitudes whose satisfaction is a sufficient condition for any predicate to qualify as a truth predicate. I will then provide a full formulation of Cognitive Command, and I will briefly describe why a discourse's satisfying this constraint justifies our thinking of the discourse as being genuinely representational. In section **(2)**, I will discuss Williamson's *original trivializing argument*. Williamson argues that *if* holding an untrue opinion constitutes a cognitive shortcoming, then any discourse that sustains the definition of a minimal truth predicate will *trivially* satisfy Cognitive Command. After presenting the argument, I will examine Wright's response. Wright argues that not only does the argument fail to apply to discourses in which the applicability of the truth predicate is *epistemically constrained*, but also that it is precisely in these discourses that showing that Cognitive Command is satisfied requires substantial philosophical work. In section **(3)**, I will discuss Shapiro and Taschek's *reductio* argument that Cognitive Command is *also* trivially satisfied by discourses whose truth predicate is epistemically constrained. After discussing Wright's *intuitionistic* response to the *reductio* argument, I will examine an argument against Wright's response and I will show why this further argument fails. I will then assess what we can learn from the failed argument about Cognitive Command. In section (4), I will conclude with a discussion of anti-realism. In particular, as Wright's intuitionistic response requires do not, I will discuss the anti-realist feature of a discourse that is registered by its failing to satisfy Cognitive Command. I will thesis about the discourse.

## (1) Minimalism and Cognitive Command

In this section I will discuss Wright's *minimalism*. First I will discuss the set of minimal platitudes whose satisfaction is a sufficient condition for any predicate to qualify as a *truth* predicate. I will then discuss the platudinous notions of *correspondence* and *representation*, and after providing provide a full formulation of Cognitive Command, and I will briefly describe why satisfying this constraint is a necessary condition for any discourse to be considered genuinely representational. In the opening chapters of "Truth and Objectivity," Wright (1992b, 28) proposes a *minimalism* about truth: he argues first that the two features of *syntax* and *discipline* of the assertoric statements of a discourse suffice to establish that the those statements are *truth-apt*, and second, that any discourse whose statements are truth-apt in this sense can sustain the definition of a *minimal truth predicate*. Which is to say, Wright maintains that if a discourse is such that the utterances of its syntactically assertoric statements are governed by norms of warranted assertibility, then it is possible to define a predicate for that discourse that satisfies a set of minimal constraints consisting of conceptual platitudes of our ordinary notion of *truth*. The basic platitudes satisfied by this *minimal* truth predicate include: (i) to assert is to present as true; (ii) any truth-apt content has a significant negation which is likewise truth-apt; (iii) to be true is to correspond to the facts; (iv) a statement may be justified without being true, and vice versa (1992b, 34). However, since the minimalist view that any discourse whose statements are truth-apt in the above sense can sustain the definition of a minimal truth predicate is intended to be neutral between any broadly realist or anti-realist conception of the discourse in question, it seems odd to include (iii) among the set of minimal constraints. For (iii) closely resembles the statement of the *correspondence theory of truth*[1], and this theory is often, though not uncontroversially, associated with the realist conception that there are objective, mind-independent states of affairs relevant to the discourse whose obtaining or failing to obtain is what renders statements in the discourse true or false.

However, as Wright (1992b, 27) points out, (iii) need not amount to such a substantive metaphysical thesis. For the motivation to include (iii) among the basic platitudes is that we would like to accommodate ordinary intuitions about the relationship between truth and correspondence. And Wright (1992b, 93) argues that we are justified in including (iii) among the minimal platitudes if we can "[demonstrate] a right to the *phrases* by which [the] intuitions about the relationship between truth and correspondence...are characteristically expressed."[2] However, the minimalist license to *speak* in this way does not settle the further question about what additional substance might be included in this thin, platudinous notion of *correspondence*.

Wright points out that a realist conception of the notion of *correspondence* essentially includes the idea that true statements in a discourse accurately represent the states of affairs distinctive to that discourse. For then it will be appropriate to suppose that the beliefs and opinions we form in the discourse are responsive or sensitive to the states of affairs that those statements are *about*, which is precisely what it is to have a realist conception of how true statements in the discourse *correspond* to the facts. So to capture a realist notion of *correspondence* we must first explain what it is for the statements of a discourse to be *genuinely* representational. To do this, we must consider more closely the notion of *representation*.

First, Wright (1992b, 91) considers the following platitude about the notion of representation:

---

[1]The correspondence theory of truth states that a statement (or proposition) "**p**" is true just in case **p** *corresponds* to the facts.

[2]Wright demonstrates that minimalists about a discourse can accommodate these intuitions and remain neutral between any more substantive realist or anti-realist metaphysics if we simply affirm both that "things are as '**p**' says they are" is an appropriate paraphrase of "'**p**' corresponds to the facts" and that "**p**" says that **p**: Assuming the disquotation schema that "**p**" is true if and only if **p**, it follows that if "**p**" says that **p**, then things are as "**p**" says they are if and only if **p**, which, by the disquotation schema gives us that "**p**" is true if and only if things are as "**p**" says they are. And this last statement is simply a paraphrase of the correspondence platitude (iii) (1992b, 25)

> (RP) If two devices function to produce representations, then if conditions are suitable and the devices function properly, then the devices will produce divergent output if and only if they are presented with divergent input.

Thus, if (RP) is a conceptual truth about the notion of *representation*, then the sense in which practitioners of a discourse are functioning in a genuinely *representational* mode is constrained by (RP). So supposing that the function of individual practitioners of a genuinely representational discourse is to produce true statements about the objective states of affairs that constitute the subject matter of that discourse, (RP) says that the opinions formed by two properly functioning practitioners of the discourse working in suitable conditions will be appropriately responsive to those states of affairs, so that their opinions will not be conflicting if the data, evidence, or other input with which they are presented is the same.

Thus, following Wright (1992b, 92-3), it seems to follow from the above platitude that we are justified in thinking of a discourse as genuinely representational just in case it is a priori that differences of opinion formulated within the discourse can be satisfactorily explained only in terms of "divergent input," or "unsuitable conditions," or "malfunction." That this condition be satisfied *a priori* is important because we want a constraint that will be satisfied by discourses in virtue of the content of their assertoric statements, and we do not want the condition to be satisfied in virtue of mere coincidence of opinion or contingent similarity among individuals' preferences. Here *malfunction* includes *any* defect in an individual's representational function—the cognitive process of belief formation engaged in by agents practicing within the discourse—including "prejudicial assessment of data or dogma" (1992b, 93). Correctable errors such as computational or inferential oversights, mistakes resulting from inattention or distraction, ignorance due to limited data, and false beliefs based on erroneous data are thus not *malfunctions* in this sense. Rather, disagreements arising under these circumstances will be attributable to either unsuitable conditions or divergent input (or both[1] ). Allowing

[1]For example, if erroneous data is collected as the result of working under unsuitable circumstances, untrue opinions formed on the basis of this data, and disputes involving these opinions, will be attributable to both unsuitable conditions and divergent input.

all of these factors to which a disagreement in a genuinely representational discourse might be attributable to fall under the heading "cognitive shortcoming," Wright (1992b, 144) reformulates the above constraint on any genuinely representational discourse as follows:

> (*Cognitive Command*) It is a priori that differences of opinion formulated within a discourse, unless excusable as a result of (i) vagueness[1] or (ii) variation in personal evidence thresholds[2] will involve something that may properly be regarded as cognitive shortcoming.

Cognitive Command is thus an additional constraint on any minimal truth predicate since its satisfaction justifies our "thinking of the judgments to which we are moved in any particular discourse as constituting the output of, as it were, a seriously representational mode of intellectual function" (1992b, 93). And while the mere availability of the minimal truth predicate in a discourse licenses our innocent talk of truths in that discourse as "representations of reality" that "correspond to the facts," etc., Wright maintains that this minimally representational feature of judgments in the discourse does not itself guarantee that disagreements in the discourse be attributable to anything worth regarding as a cognitive shortcoming. However, if a discourse does exhibit cognitive command, then it will be a priori guaranteed that all genuine disputes are attributable to such a shortcoming, be it "divergent input," "unsuitable conditions," or "malfunction." Thus, it seems, Cognitive Command is an additional feature that may or may not be exhibited in a given minimally truth-apt discourse.

[1]Wright (1992b, 144) points out that a satisfactorily precise statement of the constraint we are after should allow for the possibility that certain disputes arising from vagueness, either in the disputed statements or in the applicability to a disputed statement of the standards of warrant internal to a given discourse, may not involve a cognitive shortcoming on the part of either disputant. I mention this for completeness, though issues concerning vagueness will nowhere play a role in my discussion of Cognitive Command.

[2]For example, if I am *risk averse* in my decision making practices and you are *not*, then some disagreement we may have over what constitutes the best course of action in a given situation may not involve anything worth regarding as a cognitive shortcoming.

## (2) Williamson's Original Trivializing Argument

In this section, I will discuss Williamson's *original trivializing argument*. Williamson argues that *if* holding an untrue opinion constitutes a cognitive shortcoming, then, since contrary opinions cannot both be true, disagreements formulated within any discourse that sustains the definition of a minimal truth predicate must be attributable to the cognitive shortcoming endured by the disputant who holds a false opinion.

If sound, this argument shows that any discourse that sustains the definition of a minimal truth predicate will *trivially* satisfy Cognitive Command. It then follows from this that a discourse's exhibiting Cognitive Command does not register any substantial, realism-relevant feature of that discourse not already guaranteed by the applicability of the minimal truth predicate, and thus the constraint cannot serve to render more precise our intuition about what it is for a discourse to be genuinely representational. After presenting the argument, I will examine Wright's response and discuss one problem with this response noted by Williamson. Wright argues that not only does the argument fail to apply to discourses in which the applicability of the truth predicate is *epistemically constrained*, but that it is precisely in *these* discourses that showing that Cognitive Command is satisfied requires substantial philosophical work. Let us turn now to Williamson's primary objection.

Williamson's (1992a, 140) "original trivializing argument" can be precisely formulated as follows: Suppose that holding an untrue opinion constitutes a cognitive shortcoming. Now assume that for some discourse **D** that admits of a minimal truth predicate $true_{\mathbf{D}}$ and some statement **p** in **D**, our opinions diverge about the truth value of **p**. So I judge (or opine or believe) "**p**" and you judge (or opine or believe) "not-**p**." Now suppose that both of these divergent judgments are $true_{\mathbf{D}}$. From this it follows via the disquotation schema that **p** and not-**p**; but as this is absurd, we must reject the supposition and conclude that not both of our judgments can be $true_{\mathbf{D}}$. Thus one of our judgments must fail to be $true_{\mathbf{D}}$. As holding an untrue opinion (by hypothesis) constitutes a cognitive shortcoming, there is therefore something involved in the dispute worth regarding as a cognitive shortcoming, and thus (it is a priori that) **D** exhibits cognitive command.

As Williamson (1992a, 140) points out, if we insist that a *cognitive shortcoming* must involve the faulty operation of the faculties we employ when gathering and processing information about the world, then it is not clear that someone who endorses an untrue statement will *thereby* be guilty of any cognitive shortcoming. For we will have a sufficient amount of information to engage our judgment-forming faculties *only if* we allow defeasible evidence to stand in support of the judgments that we form. And if the evidence we have gathered is misleading, or if there is a defeater for our evidence, then the opinions we form based on that evidence may not be true. Which is to say, since "our methods of gathering information are fallible" and are not "cautious to the point of uselessness," it follows that "the healthy operation of our faculties occasionally results in the acquisition of untrue opinions without cognitive shortcoming" (1992a, 140). Williamson argues, however, that this refined notion of cognitive shortcoming robs Cognitive Command of the realism-relevant feature of a discourse that it is meant to capture. For a realist about a discourse can recognize that disputes may arise within the discourse that do not involve anything worthy of being considered a cognitive shortcoming in this more refined sense. On the other hand, Williamson concludes, without adopting some relevantly refined notion of cognitive shortcoming, the Cognitive Command constraint will not be able to resist trivialization.

Wright responds to Williamson's trivializing argument by first invoking a distinction between discourses in which the truth of statements is *epistemically constrained* and discourses in which it is not, and then by arguing that Cognitive Command is not trivially satisfied for discourses in the former class. Note that we say here that truth in a discourse **D** is *epistemically constrained* just in case the applicability of the truth predicate $true_{\mathbf{D}}$ is restricted to classes of statements (in **D**) whose truth cannot lie beyond human detection, either because these truths are knowable in principle or because it is always possible for us to gather germane evidence in their favor. Now Wright (1992a 335; 1992b 151) accepts that if the truth of statements in a given discourse is not epistemically constrained—i.e., if the truth of individual statements in that discourse is determinate and may lie beyond human detection—then that discourse will thereby satisfy Cognitive Command. For if the truth of statements in a discourse **D** is determinate and obtained independently of our ability to know or detect that these statements are true, then for any disputed statement **p** in **D**, one of the disputants involved in a disagreement about the truth of **p** will be in ignorance or error about the (determinate) situation of how matters stand with respect to the truth or falsity of **p**; and *that* disputant will be guilty of a cognitive shortcoming.[1] Wright argues, however, that matters stand differently with respect to discourses in which the truth of statements *is* epistemically constrained.

Regarding discourses where the truth of statements is epistemically constrained, Wright maintains that "holding an untrue opinion ensures...either that optimal means for determination of the truth status of the statement in question have not been implemented, or they have been misimplemented" (1992b, 336). For if the truth of statements in the discourse cannot altogether lie beyond human recognition, then fully exercising the appropriate capacities relevant to determining the truth value of statements in the discourse would necessarily lead one to form true opinions. Thus, Wright argues, for discourses where truth is epistemically constrained, the claim that someone is guilty of a cognitive shortcoming *in virtue of* holding an untrue opinion is justified if and only if we can be guaranteed that there is some "in principle *independently appreciable* shortcoming," some defect or impropriety, "in the grounding of that opinion" (1992b, 336). Moreover, Wright points out, providing this guarantee will require substantial philosophical work, and anyone who thinks that a given discourse fails to exert Cognitive Command will believe that no such guarantee can be had. Thus, Wright concludes that Cognitive Command is not "open to trivialization" in the way threatened by Williamson's argument, since "claiming Cognitive Command for a discourse is a commitment to serious philosophy, and is defeasible by a failure to discharge that commitment" (1992b, 151).

However, it is not clear Wright's response to Williamson's original trivializing argument successfully blocks Williamson's. In fact, as Williamson (1992a, 140) points out, Wright's response to the trivializing argument seems "off the point." Recall that Wright responds to the objection by arguing that an effective defense of the

[1]Note, similarly, that if the Law of Excluded Middle (LEM) holds for a discourse **D**, then that discourse will also satisfy Cognitive Command. For suppose that for some statement **p** in **D**, disputant A judges **p** and disputant B judges not-**p**. By LEM (in **D**), we know that (i) **p** ∨ not-**p** is $true_{\mathbf{D}}$. But if **p** is $true_{\mathbf{D}}$, then B, being in error about the truth of **p**, is guilty of a cognitive shortcoming; and if not-**p** is $true_{\mathbf{D}}$, then A is similarly guilty of a cognitive shortcoming. So, by disjunction elimination, we can conclude from (i) that the dispute involves something worth regarding as a cognitive shortcoming and thus that **D** satisfies Cognitive Command.

claim that Cognitive Command is met for a discourse must show either (i) that truth in the discourse is not epistemically constrained or (ii) that truth in the discourse is epistemically constrained and disagreements formulated within the discourse involve a defect or impropriety in the manner in which one of the disputants determined the truth value of the disputed statement. Thus, Wright points out, a discourse cannot *trivially* exert Cognitive Command since both possible defenses of the claim that the constraint is met require substantial philosophical work. However, Williamson's objection was *not* that if a discourse satisfies Cognitive Command, then it must do so for the *trivial* reason that disagreements formulated within the discourse are attributable to one or other of the disputants being in ignorance or error about the truth value of the disputed statement. Rather, the objection was that Cognitive Command is *itself* trivially satisfied. And while it may be correct for Wright to point out that one's holding an untrue opinion in a discourse where truth is epistemically constrained will involve a failure to fully exercise one's capacities for determining the truth value of the opined statement, this does not prevent the mere *holding of an untrue opinion* itself from already constituting a cognitive shortcoming. Thus, Williamson maintains, without modifying the notion of *cognitive shortcoming* so as to block the original trivializing argument, the trivializing argument shows that every discourse sustaining the definition of a minimal truth predicate will trivially satisfy Cognitive Command. I will not here further assess the force of Williamson's trivializing argument or the correctness of Wright's response. Rather, I will now turn directly to the *reductio* argument of Shapiro and Taschek, for, as we shall see, Wright's *intuitionistic* response to reductio argument serves equally well as a response to the trivializing argument raised by Williamson.

## (3) Shapiro and Taschek's *Reductio* and Wright's Intuitionistic Response

In this section, I will discuss Shapiro and Taschek's *reductio* argument that Cognitive Command is *also* trivially satisfied by discourses whose truth predicate is epistemically constrained. After discussing Wright's *intuitionistic* response to the *reductio* argument, I will examine an argument against Wright's response, and I will show why this further argument fails. I will then assess what we can learn from the failed argument about Cognitive Command. Finally, as Wright's *intuitionistic* response requires us to revise the original distinction between discourses that satisfy Cognitive Command and those that do not, I will conclude by discussing the *anti-realist* feature of a discourse that is registered by its failing to satisfy Cognitive Command.

Taschek and Shapiro (1996, 85) present the following variation of Williamson's argument to show that Cognitive Command is trivially satisfied for discourses where truth is epistemically constrained: Suppose that for some discourse **D** that admits of a minimal truth predicate $true_{\mathbf{D}}$, this notion of truth operative in **D** is epistemically constrained, so that for any statement **s** in **D**, if **s** (is true), then **s** is knowable. Now assume that the **D** fails to exert cognitive command; so for some statement **p** in **D**, subject *A* believes (opines or judges) **p** and subject *B* believes (opines or judges) not-**p**, yet neither disputant has implemented the means for determining the truth value of **p** in a less than optimal way. But now assume **p**. It follows from this assumption and the condition of epistemic constraint that **p** is knowable. But then *B* believes the negation of a statement that is both true and knowable, and thus *B* suffers a cognitive shortcoming, contradicting the assumption that cognitive command fails for **D**. So we must reject the assumption and infer not-**p**. But then a symmetric argument shows that *A* suffers a cognitive shortcoming, since *A* believes something whose negation is both true and knowable, again contradicting the fact that **D** exhibits cognitive command. Thus we must reject our initial assumption and conclude that **D** does not fail to exert cognitive command. And since this *reductio* argument holds for any discourse in which the truth of statements is epistemically constrained, we can conclude *a priori* that any discourse where truth is epistemically constrained will trivially satisfy Cognitive Command.

It is important to note that the cognitive shortcomings attributed to the disputants throughout the course of the above argument are in fact shortcomings that are appreciable *independently* of the mere fact of disagreement. For if **p** is true and *B* believes not-**p**, the cognitive shortcoming endured by *B* does not involve any mention of his participating in a dispute with *A*. The result is analogous if not-**p** is true and *A* believes **p**. So if Shapiro and Taschek's objection stands, then it, together with Williamson's original trivializing argument, will show that Cognitive Command is in fact satisfied by *any* minimally truth-apt discourse. Let us now examine how Wright responds to Shapiro and Taschek's objection.

Wright responds to the objection by insisting that we should reject the final, *classical* inference[1] of the argument, which purports to establish that cognitive command is satisfied by a discourse since, in that discourse, Cognitive Command cannot be said to fail. To motivate this response, Wright argues that there is good reason to endorse the distinction between a discourse's *not failing* to satisfy cognitive command and our having an a priori guarantee that the condition is met.

Wright (1992a, 334-5) clearly acknowledges that the *reductio* argument forces us to deny that no cognitive shortcoming is involved in disputes between individuals holding contradictory beliefs in a discourse where truth is epistemically constrained. However, Wright argues that this conclusion simply means that "there can be no mandate for the claim that *no* shortcoming is involved," and thus that it is "permissible to think that cognitive shortcoming is involved in the dispute" (1992a, 334). And he argues that this alone is not sufficient to guarantee that cognitive command is satisfied. For the satisfaction of cognitive command requires explicitly that we have an *a priori guarantee* that no cognitive shortcoming is involved in the dispute, and its being permissible to assert as much does not give us the requisite guarantee, provided that we do not allow a "double negation elimination step on the immediate conclusion of the *reductio*" (334). Thus, Wright concludes, the relevant distinction between discourses that exert Cognitive Command and those that do not is a distinction between "those wherein disagreements must, subject to the other [proviso] conditions, be ascribed to cognitive shortcoming and those which make possible disagreements for which such an ascription is not mandated" (1992a, 335). In other words, a discourse that *does not* exert Cognitive Command is not one in which it is possible to formulate disagreements that involve nothing worth regarding as a cognitive shortcoming, since this is never the case, but rather it is one in which it is possible to formulate disagreements with respect to which there is no *mandate* for the claim that some cognitive shortcoming is involved. And, Wright concludes, "that is exactly as it should be if conflicting opinions are permissible" (1992a, 335).I will now consider a potential objection to Wright's *intuitionistic* response, and I will examine why the objection fails. Discussing this failed objection is interesting

[1]This inference is an instance of *double negation elimination*, which is not *intuitionistically* valid. Note, however, that all of the steps in the *reductio* argument *are* intuitionistically valid, except for the final step.

insofar as it makes more precise Wright's distinction between discourses that do not fail to satisfy Cognitive Command and those that actually succeed in meeting the constraint. A further upshot of the failed objection is then that it helps to indicate what must be accomplished in order to establish that a discourse does not succeed in satisfying Cognitive Command.

The argument I will consider is that the double negation elimination step of the *reductio* is in fact justified in the current context due to specific considerations relevant to the notion of *cognitive shortcoming* (and not to any more general support for the validity of this inference rule). First, I will present the argument[1], and then I will examine why it fails.

The mistaken argument assumes that if a cognitive shortcoming is involved in a disagreement formulated in an epistemically constrained discourse, then this cognitive shortcoming is something that cannot lie beyond our detection, and concludes from this assumption that there is a provable lack of counterexamples to Cognitive Command for discourses satisfying epistemic constraint. The argument, I claim, is valid, but the assumption of the argument is false. If, however, the argument were sound, then it would follow that not only by the *reductio* could we *not* say of any discourse that there is the possibility of disagreements formulated in the discourse that involve nothing worth considering a cognitive shortcoming, but furthermore, we *could* say of the discourse that disagreements in which there is nothing worth regarding as a cognitive shortcoming are not, in fact, possible. The argument runs as follows:

Given any discourse **D** that admits of a minimal truth predicate $true_{\mathbf{D}}$ and is such that truth in **D** is epistemically constrained, the *reductio* argument shows that **D** cannot be said to fail to satisfy Cognitive Command. Let a counterexample in **D** to

[1]The motivation for the objection comes from the theorem in Model Theory that independent $\Pi^0_1$ sentences formulated in the language of formal number theory (say **PA**) using decidable predicates are true in the standard model. For suppose that some $\Pi^0_1$ sentence $\forall x\phi(x)$ is independent of **PA**, where $\phi()$ is a decidable predicate. Then there are no counterexamples to $\forall x\phi(x)$ in the standard model: if there were a counterexample, then we could prove that it is a counterexample, contradicting the independence of the statement $\forall x\phi(x)$. That is, suppose that n is a counterexample to $\forall x\phi(x)$ in the standard model of **PA**. Then $\neg\phi(n)$ is true in this model, and since $\phi()$ is a decidable predicate, we can prove $\neg\phi(n)$, and thus we can prove $\neg\forall x\phi(x)$, contradicting the fact that $\neg\forall x\phi(x)$ is not provable, since $\forall x\phi(x)$ is independent of **PA**. So there are no counterexamples to $\forall x\phi(x)$, and thus this statement is true in the standard model.

Cognitive Command be any statement **p** in **D** such that a disagreement can occur (in **D**) about the truth value of **p** without involving anything worth considering a cognitive shortcoming. Now suppose there were such a counterexample in **D** to Cognitive Command, say $\mathbf{p}_0$ . Then we can imagine two disputants *A* and *B*, such that *A* believes $\mathbf{p}_0$ , *B* believes not-$\mathbf{p}_0$ , and neither *A* nor *B* suffers anything worth considering a cognitive shortcoming.[1] But then if there is nothing involved in the dispute (i.e., in how the subjects determined the truth value of $\mathbf{p}_0$) worth considering a cognitive shortcoming, then we can *prove* that that no such defect is involved. That is, we can examine the process or method of reasoning employed by each of the disputants *A* and *B* to determine the truth value of $\mathbf{p}_0$ and prove that there is nothing involved in these processes worth considering a cognitive shortcoming. For we can simply examine each individual's *evidence* for his or her judgment, and the *method of reasoning* employed to justifyhis or her opinion based on this evidence; and as the evidence and argument grounding an individual's judgment are finite, we can simply prove that there are no considerations involved in the grounding of either individual's opinion that mandate withdrawal of his or her respective view. Thus, if no cognitive shortcoming is involved in the dispute, we can *prov*e that no such shortcoming is involved, thus *proving* that $\mathbf{p}_0$ is a counterexample to Cognitive Command. But this proves that **D** fails to exhibit cognitive command, contradicting the result of the reductio argument. Therefore, we must conclude that there are no counterexamples in **D** to Cognitive Command.

However, the argument fails for the following simple reason: it ignores the fact that beliefs formed under *unsuitable conditions*, in addition to those formed with *faulty input* (i.e., evidential support) or due to a malfunction (i.e., defect in method of reasoning), also require attributions of cognitive shortcoming. For while our justifications for our claims (our evidence and method of reasoning grounding those opinions) are surveyable, it is not clear that there will not be some unsuitable feature of the *conditions* under which the opinions were formed that escapes our notice.[2]

[1]That is, in the sense of *cognitive shortcoming* relevant to a discourse where truth is epistemically constrained, neither *A* nor *B* is guilty of a suboptimal implementation of the means for determining the truth value of the disputed statement $\mathbf{p}_0$.

[2]Or, to be more precise, though we cannot assert that there is such an unsuitable feature (i.e., we cannot assert ∃x...), we will not *thereby* be able to assert that there is no such relevant unsuitable feature (i.e., we cannot *thereby* assert ∀x¬...)

I will not here address the difficult question of what is included in the conditions under which our beliefs are formed or what is required for these conditions to be *suitable* or *unsuitable* in a given context. But supposing, in a given disagreement, that there may be some unsuitable feature of the conditions under which the opinions were formed that escapes our notice, then our ability to provide a positive guarantee that cognitive shortcoming is involved is undermined. For if we can have no a priori guarantee that we would detect such an unsuitable feature, then we can have no a priori guarantee that something worth considering a cognitive shortcoming must be involved in such disputes. And without such a guarantee, Cognitive Command is not satisfied.

Thus, Wright's distinction between discourses that *do not fail* to satisfy Cognitive Command and those that actually succeed in meeting the constraint seems to survive close scrutiny. Moreover, we have seen from the above argument that if we lack a guarantee that Cognitive Command is satisfied for a given discourse, it is because we lack a guarantee that there is no unsuitable feature of the conditions under which participants in a given disagreement formed the opinions that give rise to the dispute. Such an unsuitable feature will be either a feature of the conditions under which an individual gathered evidence for her claim or a feature of the conditions under which she assessed this evidence, and will not involve any explicit, observable fact about her justification for her claim or any failing in her actual assessment of the evidence. Therefore, a guarantee that Cognitive Command is satisfied by a discourse must consist in a guarantee that disputes formulated within the discourse are not attributable to unsuitable conditions.

## (4) Anti-Realism: *Non-Representational* and *Relativist* Discourses

In this section, I will conclude with a discussion of *anti-realism*. In particular, as Wright's *intuitionistic* response to the *reductio* argument requires us to revise the original distinction between discourses that satisfy Cognitive Command and those that do not, I will discuss the *anti-realist* feature of a discourse that is registered by a discourse's failing to satisfy Cognitive Command. I will begin by pointing out that a discourse's satisfying Cognitive Command *still* serves to justify our thinking of the discourse as *genuinely representational*. I will then note the worry that a discourse's *failing* to satisfy Cognitive Command might register no anti-realist feature of a discourse at all, and I will argue that this worry is misguided. Finally, I will conclude by arguing that even in light of the modified distinction forced by Wright's intuitionistic response, claiming that a discourse fails to satisfy Cognitive Command is one way of articulating a relativist thesis about the discourse.

First, it is important to observe that a discourse's satisfying Cognitive Command *still* justifies our thinking of the discourse as *genuinely representational*. For if we are in fact in possession of a *guarantee* that differences of opinion formulated in a discourse will be attributable to some cognitive shortcoming, then we are justified in both (i) thinking of judgments expressible in the discourse as the output of a "seriously representational mode of intellectual function" and (ii) thinking of *true* judgments in the discourse as accurate representations of the self-standing facts that constitute its sub ject matter (1992b, 93). Moreover, this realism-relevant feature of a discourse is not already guaranteed by its sustaining the definition of a minimal truth predicate, and thus the original intuition behind the formulation of the Cognitive Command constraint is at least partially respected.

However, since we can no longer characterize failures to satisfy Cognitive Command in terms of the genuine possibility of *cognitively blameless* disagreement—disagreement involving *nothin*g worth regarding as a cognitive error or shortcoming—a discourse's *failing* to satisfy Cognitive Command seems to no longer justify our thinking of that discourse as *non-representational* or as lacking any self-standing facts that assertions in the discourse serve to represent. Rather, it appears that a discourse's failing to satisfy Cognitive Command merely means that we have no *guarantee* that the discourse is representational, and this, admittedly, is less strong than the original intuition that we wanted to respect. Moreover, since it is logically consistent (though perhaps unreasonable or incoherent) to assert that a genuinely representational discourse can fail to satisfy the constraint—since this conjunction is just the pair of assertions that (i) the discourse is genuinely representational but (ii) that we lack a guarantee that this is so—one might worry that the possibility that a discourse's failing to satisfy Cognitive Command registers any anti-realist feature of the discourse *at all* is thereby undermined.

There are two things that can be said to assuage this initial worry that a discourse's failing to satisfy Cognitive Command registers no anti-realist feature of the discourse. First, if one accepts the "*general* rule" that the "prephilosophical or 'default' stance about a discourse should be one of parsimony"' and thus that "realism must be earned," then our lacking a guarantee that a discourse satisfies Cognitive Command suggests that we should retain this default stance and take care to avoid the unjustified metaphysical assumption that the discourse is genuinely representational (1992b, 149). Second, it is not the case that an anti-realist who believes a discourse fails to satisfy Cognitive Command is in no position to *deny* that the discourse is genuinely representational, nor is it the case that he or she must merely insist to a realist opponent that they should both remain responsibly agnostic about this point. Rather, the anti-realist can argue that the fact that the discourse fails to satisfy Cognitive Command suggests that the metaphysical pretensions of a realist conception of the discourse are utterly unjustified, even if he or she does not take him or herself to have any *proof* that this realist conception is false. Clearly, such an anti-realist will have to employ additional resources beyond the fact a discourse fails to satisfy Cognitive Command if he or she she wishes to guarantee that the discourse is non-representational. Nonetheless, it is still possible for an anti-realist to coherently deny that a discourse is genuinely representational. Quoting Wright (2006, 51):

> The dialectical situation is, in fact, exactly analogous to that in the philosophy of mathematics, where the intuitionist may quite coherently—if he wishes—deny the Platonist metaphysics of a crystalline world of determinate mathematical structures, potentially conferring truth and falsity upon our mathematical statements in ways transcending all possibility of proof. That denial commits him to denying that Excluded Middle holds of necessity for reasons connected with that metaphysics. But it does not commit him to denying Excluded Middle itself, still less any instance of it. Rather, in the absence of justification for the principle of any other kind, he simply regards it as unacceptable.

Similarly, in our case, the anti-realist who denies that a discourse is representational can thus deny that there are any self-standing facts that provide an a priori guarantee that disputes formulated in the discourse will involve something worth regarding as a cognitive shortcoming. But his or her denial that a discourse is representational will still not allow him or her to *deny* that disputes formulated in the discourse involve anything worth regarding as a cognitive shortcoming; nor will it allow her to suppose that the discourse contains any actual counterexamples to Cognitive Command. For as we have seen by Williamson's *trivializing argument*, these claims are in conflict with "the most basic constitutive principles regarding truth and negation" (2003, 498). Rather, in the absence of a guarantee that disputes formulated in the discourse involve anything worth regarding as a cognitive shortcoming, the anti-realist merely regards Cognitive Command as unacceptable.

However, even if an anti-realist can argue that a discourse is non-representational, there is still no provision offered by the current framework for her to argue that genuine disagreements can arise in a discourse in which the views of the two parties to the dispute are both *correct*. And Wright (2006, 49-50) recognizes it may be felt as a limitation of this framework that it cannot accommodate the intuition that genuine disagreements may arise in discourse without involving any cognitive shortcomings or *faults*. Moreover, Wright recognizes that this feeling may be especially strong if one insists that *true relativism* about a discourse should be characterized as the view that the discourse contains genuine fault-free disagreements. For, in that case, the frame-work under discussion does not allow for the coherent expression of a true relativistic thesis.

However, I do not believe that this is the conclusion we should draw. Rather, I want to here conclude by showing that there is room in the current framework to argue that a discourse's failing to satisfy Cognitive Command justifies a *modified* relativism about the discourse. To establish this claim, one could first argue that the proper lesson of Williamson's *original trivializing argument* is that it is not possible to coherently attach any content to the idea of a genuine disagreement in which the *inconsistent* opinions of the two parties to the dispute are nonetheless both correct (2006, 50-51).[1] This would show that the *true relativist* thesis is

[1]In other attempts to make sense of this relativistic thesis or account for the possibility of this type of fault-free disagreement, such as MacFarlane's (2005) or Wright's (2006, 52-59), it is somewhat less clear that the disagreements formulated in relativistic discourses remain *genuine*—i.e., are disagreements about the truth-value of an individual proposition. For in general if one introduces a relativistic parameter into the propositions

accordingly misconceived and would suggest to the anti-realist that she should instead adopt the *modified* relativist view that differences of opinion can arise in a discourse for which there is *no mandate* to attribute fault. This modified relativism seems to be a perfectly adequate and natural view, especially if the goal of the anti-realist is to express the idea that a discourse may contain two logically inconsistent views that are both nonetheless *permissible*. Finally, one could argue via Wright's intuitionistically valid distinction that a mandate to attribute fault in a dispute requires more than just our being in a position to exclude the possibility that both parties to the dispute are correct. Rather, one could insist that a *mandate* to attribute fault in a dispute requires being in possession of an actual *guarantee* that some cognitive shortcoming is involved in the disagreement. From this it would follow that (i) disagreements formulated in a discourse involve a mandate to attribute fault to one or other of the disputants *only if* that discourse satisfies Cognitive Command, and thus that (ii) a discourse's failing to satisfy Cognitive Command is a sufficient condition for the claim that the discourse may contain pairs of logically inconsistent views that are both nonetheless *permissible*. If these arguments are correct, then they establish the claim that once we take the force of Williamson's *trivializing argument* seriously, a proper *relativism* about a discourse may be formulated as the view that although the discourse is minimally truth-apt, it fails to satisfy Cognitive Command.

disagreements formulated in relativistic discourses remain *genuine*—i.e., are disagreements about the truth-value of an individual proposition. For in general if one introduces a relativistic parameter into the propositions expressed by utterances in the relativistic discourse—i.e., if one insists that an utterance of the sentence "**s**" serves to express the proposition "**s** is true *relative to*..."—then it is not immediately clear that disputes over the truth value of **s** will be genuine in the sense described. For one disputant will be claiming that **s** is true *relative to*..., the other disputant will be claiming that **s** is not true *relative to*..., and the relativistic parameters in both propositions might not coincide. Thus, if one insists that in discussing relativism one should be concerned with genuine disputes in the above sense, then this framework under discussion clearly recommends itself to help articulate the relativistic thesis.

## Works Cited

MacFarlane, John. “Making Sense of Relative Truth.” In *Proceedings of the Aristotelian Society*, 105 (2005) pp. 321-39.

Shapiro, Stewart, and William W. Taschek. “Intuitionism, Pluralism, and Cognitive Command.” In *The Journal of Philosophy*, Vol. 93, No. 2 (1996) pp. 74-88. 21

Williamson, Timothy. “Critical Notices: Truth and Objectivity.” In *International Journal of Philosophical Studies* (1992a) pp. 130-144.

Wright, Crispin, “Critical Notices: Realism, Pure and Simple? A Reply to Timothy Williamson.” In *International Journal of Philosophical Studies* (1992a) pp. 327-341.

———. *Truth and Objectivity*. Cambridge : Harvard University Press, 1992b.

———. *Saving The Differences: Essays on Themes from Truth and Objectivity*. Cambridge: Harvard University Press, 2003.

———. “Intuitionism, Realism, Relativism, and Rhubarb.” Forthcoming in Greenough and Lynch, editors, *Truth and Realism*, pp. 38-60. Oxford: Oxford University Press, 2066.

# Interview with Ned Block, New York University

Professor Block has written extensively on a number of topics in the philosophy of mind, from consciousness to cognitive science, and is particularly known for his work on Functionalism. Many of his papers are collected in *Consciousness, Functionalism, and Representation* (2007). This interview was conducted in New Haven on March 6th, 2007.

**The Yale Philosophy Review:** *You're a professor of both philosophy and psychology. I'm curious what you see as the connection between these two fields and, more broadly, what you think science has to contribute to philosophy or vice versa. And maybe even what role remains for the methods of traditional philosophy in contemporary philosophy of mind.*

**Ned Block:** Okay, so from the point of view of a philosopher I think that a lot of the most interesting issues in philosophy of mind are the ones that are partly empirical. That's certainly true for consciousness, where the work that's being done in neuroscience is revolutionizing our way of thinking about consciousness. I think this has been obvious for a really long time. I don't know if you guys have read Tom Nagel's 1971 paper on brain bisection, called "Brain Bisection and Unity of Consciousness." He looks at these cases of people whose two hemispheres have been surgically separated to keep epilepsy from moving from one hemisphere to the other. He argues rather persuasively that this neuroscience evidence creates problems for our concept of a person—so there's a philosophical issue where neuroscience changes everything in terms of the way you have to think about it. There's another way in which the sciences of the mind impinge, which is that they allow you to clarify your concepts, so in my talk today I was talking about the relation between consciousness and accessibility, which, in one way or another, just has been involved in almost everything that philosophers have done on consciousness. Work in the sciences of the mind suggests that there are a number of different notions of accessibility that we have to distinguish and they tell us something about their properties, and that's certainly very important in thinking about the relationship between consciousness and accessibility. In other areas like concepts, for example, you know, some appreciation of the psychological linguistics work on concepts helps us think about not only traditional philosophical problems but about conceptual issues related to traditional problems that come up in some new way when you look at the sciences.

**YPR:** *Do you think there's any room today for philosophy devoid of empirical evidence?*

**NB:** I have always had two streams of work: one relatively "armchair" and one relatively empirical. So for So for example, I've just done an article on Wittgenstein which looks at inverted spectrum issues and is relatively unempirical. I think that it's really a matter of degree because there's always some empirical relevance—as soon as you start talking about, for example, anything to do with color, you have to look at the empirical work on color. That has to at least be in the background, and I've written plenty of things that are pretty *a priori* but not really completely *a priori*. The parts of it that are completely *a priori* are much less interesting to me than the parts that aren't so *a priori*.

**YPR:** *To jump to a related question, then, what do you think of recent work in what is called "experimental philosophy"?*

**NB:** I'm favorably disposed in principle, but it depends what you mean by experimental philosophy. So, one thing that people call experimental philosophy is, you go and ask people questions about their intuitions about knowledge, for example. Philosophers say in order for you to know something, it's necessary that it be true that you believe it, and that you're justified in believing it. But people often say they knew something that turned out to be false that they were totally certain about. The problem with questionnaires is that philosophers are trained in these intuitions, and when you ask people you have to sort of train them to get them to understand the difference, for example, between a semantic intuition and a pragmatic intuition. I don't know if that kind of thing is really going to work, since I haven't really studied it. But some very smart people are engaged in it. Another thing people sometimes include in “experimental philosophy” is an experimental approach to “moral psychology,” where you look at what answers people give to moral questions and what happens in their brains when they are answering. For example, Josh Green and Mark Hauser at Harvard are doing this kind of work. Josh has a PhD in philosophy and he has a job in a psychology department. Now, there I think you can find some things that are very relevant to how people should think about ethics. The machinery, the mechanisms that generate our moral intuitions are not irrelevant to the rationality of those moral intuitions. So I think that’s potentially very exciting work.

**YPR:** *Cognitive science has sometimes been criticized as having problematic theoretical grounds. Some have called it too empirical, and some have called it too philosophical. I’m wondering how you would reply to that and whether you think there’s any truth to that. Is there any room for improvement in the theoretical underpinnings of cognitive science, or do you think these criticisms are mostly unfounded?*

**NB:** I would distinguish between criticizing cognitive science as a field and the work done by individual cognitive scientists. Individual cognitive scientists, like individual philosophers, are often confused. But, as a field, I think the field is well grounded enough. It’s still a field that’s moving from what Kuhn would have called a pre-paradigm stage into a paradigm stage. So there’s still a substantial amount of disagreement among psychologists about exactly what the right methods are and a lot to be done in clarifying assumptions. But there’s all kinds of wonderful, interesting stuff in cognitive science and I think it’s doing very well as sciences go.

**YPR:** *Do you think that at some point people will be trained as cognitive scientists, versus, as far as I understand it, people from different fields coming together and contributing?*

**NB:** When cognitive science first took on that name in the seventies there was a lot of discussion of whether there should be cognitive science departments. Most people with PhDs in cognitive science couldn't get jobs except in cognitive science departments. It's a general problem about creating a new department in anything, that it's very difficult for people to get jobs outside of the field in which they earned their PhD. So it's hard to get them started; there's a bootstrapping problem, you have to somehow get off the ground. There are a couple cognitive science departments, not many, but there won't be many PhD programs until there are more.

**YPR:** *I have read that you've expressed optimism as to the possibility of closing the explanatory gap between mind and body. What would you consider the strongest reply to someone who claimed it's an unsolvable problem?*

**NB:** I don't have any positive argument that it's solvable, but neither do I think the arguments that it's unclosable are any good. So I think it remains an open issue. If it is to be closed, it's only going to be via ideas that arise in finding solutions to "easy" problems, so investigating problems is the methodology I would recommend. The explanatory problem is the "hard" problem; the "easy" problems are problems to do with functions of consciousness. We're certainly learning lots more about various features of consciousness, its functions, how it relates to other parts of the mind. If we're ever going to close the gap it's going to be that way. I think there are ideas we need to have, but in order to get those ideas we have to work on these "easy" problems.

**YPR:** *Can you give a rough sketch of the kinds of advances or shifts in ideas that would be required to reach that point, or is it impossible to say?*

**NB:** I think the best we have is analogies with areas like life, for example. An often-made analogy where people just didn't see any way science could explain life, and then what happened was something quite unanticipated, is that all the functions that are involved in life could be given a physical basis, the most important of which was the physical basis of the genetic code. I don't think it's very easy to predict exactly what form the breakthroughs are going to be in before they happen.

**YPR:** *To what extent do you think the problems of consciousness are metaphysical* *problems? They're usually thought of as having ramifications for questions of personal identity and the ontological status of mind—are there any other points at which the philosophy of mind runs into metaphysics?*

**NB:** Well, the mind-body problem is a metaphysical problem. What is the essence of mind, is it material, is it functional, soul-stuff? I think there is a metaphysical issue of what makes all pains pains, as well as an ontological issue about the material or immaterial status of pain; that's one of the central issues in philosophy of mind. I think it's filled with metaphysics. I have been less interested in the epistemology of mind although that is now becoming an active area of research.

**YPR:** *Conversely, what responsibility, if any, do you think metaphysics has to engage with problems of mind?*

**NB:** The fields do overlap, but someone can study issues in metaphysics that don't impinge on the mind: whether there are numbers, for example. On the other hand, an issue like vagueness—there I think there's more contact. I think some issues in metaphysics are outside of the metaphysical problems of mind, and some are relevant to philosophy of mind.

**YPR:** *To which historical philosophy would you say contemporary philosophy of mind owes the most? And to which do you see yourself as owing the most?*

**NB:** Leibniz famously had this image of mill, which is relevant to issues about functionalism. If we were inside of a mill we wouldn't see what it was from the inside. The analogy is used today to counter anti-functionalist intuitions. Of course Wittgenstein is very important to anything do to with the mind, but I can't see that there's any historical philosopher whose work is really totally central—except maybe Descartes. Usually we discover the views of great philosophers after we've already been thinking about them independently. Victor Caston had a classic paper on Aristotle trying to argue that he holds a very popular view on the nature of awareness, [that it's] a reflexive representation. But I don't think anyone would have ascribed that to Aristotle until they'd come upon it themselves already.

**YPR:** *When you refer to Wittgenstein are you referring to his early or late work?*

**NB:** Late. I mean the late work on consciousness and behavior and that kind of thing. I myself think of it as the wrong approach, but you have to be influenced by it.

**YPR:** *Where do you stand in the debate as to the usefulness or the validity of metaphysics as a field of inquiry? And to what extent does your view here inform your understanding of philosophy of mind?*

**NB:** I talk a lot on the mind-body problem. That's the problem of what's the right metaphysical picture of mind. I think the leading contenders are physicalism of one sort or another, functionalism, which takes the computer model seriously, and dualism. And that's a really important debate. Is the computer metaphor right? Is it more neurological or biological? Or are both of those wrong and should we be thinking more in terms of religion? That's a major debate that intersects with all kinds of interesting issues, and it's metaphysics. So I'm very pro-metaphysics.

**YPR:** *How would you weigh* a priori *reasoning versus experimental data in considering metaphysical questions? What percentages would you assign?*

**NB:** I don't think you can assign a percentage. Empirical data has to be interpreted. It doesn't mean something all by itself. You need at least relatively *a priori* reasoning in order to interpret the empirical data. I'm not a big believer in *a priori* truth, so I have questions as to whether there's any really purely *a priori* reasoning, but still relatively a priori reasoning is certainly very important in the sciences.

**YPR:** *Do you think of yourself as primarily a philosopher as opposed to a psychologist?*

**NB:** Oh yes, definitely. I'm a philosopher.

**YPR:** *Why?*

**NB:** I guess what I find fun is thinking about these more theoretical questions. I've never done an experiment. I like looking at the results that psychologists get and trying to figure out what they mean, so that makes me more of a philosopher. But I think the border between philosophy of mind and theoretical psychology and theoretical cognitive neuroscience is very hard to draw; there isn't really much of a border. I think there are a lot of things that could be done in any number of different departments. A lot of scientists do things that involve a large philosophical element. You have a great psychology department here, and there are a lot of people in it who are pretty philosophically inclined.

**YPR:** *Do you think that philosophy of mind will continue to make the same progress that is has for the last forty or fifty years?*

**NB:** Prediction is risky, but because it gets its energy from the sciences of the mind and I don't see any sign of that going away, I think, yes, it's likely to. Certainly my entire career has been getting steadily more interesting. Fields do come and go in philosophy: for example, when I was in graduate school, epistemology was very unpopular. But in a field largely driven by an adjacent field that's funded by grants and certainly is going to keep going, I think that is going to ensure that philosophy of mind will continue to be at least as popular [as it currently is].

**YPR:** *Would you say that the ultimate goal of philosophy of mind is to eventually become a science?*

**NB:** No, I think that there will always be conceptual issues in the sciences of the mind that have something to do with traditional philosophical concerns. Philosophers will always be interested in this.

**YPR:** *What do you think the role of intuitions and folk psychology is?*

**NB:** In the case of consciousness, the place to start is what we know about our conscious states from our first person point of view. Now some of those things we think we know we may not actually know, so you have to be somewhat skeptical.

I take that first-person point of view very seriously. I think you should start with those intuitions and then ask yourself where they come from and see if you can go beyond them. I think it's important not to just rely on intuitions without asking yourself why you have those intuitions. Some intuitions reflect concepts and tell you something about the concepts; others just reflect prejudice or outdated science.

# Interview with Robert Pippin, University of Chicago

Robert Pippin is the chair of the Committee on Social Thought at the University of Chicago, where he is the Evelyn Stefansson Nef Distinguished Service Professor. Most well-known for his work in German idealism, he is the author of *Hegel's Idealism: The Satisfactions of Self-Consciousness* (1989), *Modernism as a Philosophical Problem: On the Dissatisfactions of European High Culture* (1991), and *The Persistence of Subjectivity: On the Kantian Aftermath* (2005), among other works. We felt that his insights on a number of topics both philosophical and non-philosophical easily warranted the relatively unedited version you see here. Even so, our discussion with him far outlasted an hour of tape, so a good deal has still been left out. The interview was conducted in Linsley-Chittenden Hall of Yale University on March 28th, 2007.

**Yale Philosophy Review:** *What got you interested in German idealism?*

**Robert Pippin:** I could probably use up all of your tape; it's a long story. In college I was not a philosophy major. I was a literature major, but I had studied a lot of philosophy with some people I respected a lot and thought quite a lot of. I got interested in two problems. On was—in literature especially—the problem of self-knowledge, and [the other was] the problem of history: why (particularly normative) institutions changed so radically. And that led me to a general interest in various theories of modernity. And I was working this all out in one way or another by doing literary criticism, worrying about the place of literary products within particular ages and the relations of the ages to each another and so forth, and by the end of the day, it just seemed to me that I could do what I was interested in, work on a set of problems and their historical inflections, more easily in philosophy departments. But I got into philosophy by being exposed mostly to the European tradition. I was an undergraduate at Trinity College, and most of the people I took courses with were in the history of philosophy and European philosophy. I didn't really have a particularly good education in Anglophone philosophy at that point, so I was advised to go to Penn State, and I did, and it had a very eccentric department, full of Heidegger experts and Levinas people and Straussians. It was just an incredible potpourri of people in the early 1970s. And there, again, most of the stuff I was studying came out of the European tradition, and within that tradition, the decisive figure is Kant, and, because of the shift into a more historicized conception of philosophy in the European tradition, Hegel. So I was thoroughly immersed in that tradition by virtue of where I was going to school, what teachers I had, what I was interested in, and what advice I was given. The other decisive thing to remark about that background is that, in my second year of graduate school, the chair of the department, who had been a student of Wilfred Sellars, persuaded him to fly once a week from Pittsburgh to State College to teach a Kant seminar. I was just completely blown away by it. He was the smartest person I'd ever seen, and he had the most interesting things to say about a variety of things in the history of philosophy, and as far as I could tell, far more interesting than the so-called historians of philosophy that I had been studying with. So Sellars opened another door, that the issues I was studying were not hermetically sealed within the European tradition, and that with the right language and the right education, they were topics that were broached by the best minds in the Anglophone tradition as well. That I count as a great benefit; that experience was a way of seeing how one might think about these issues not simply within the so-called European tradition.

**YPR:** *Hegel has enjoyed a rise in Anglophone attention since the sixties, one that is still going on. This ranges from attention to his influence on Marx and Marxism to discussions, more recently, of the Other in political theory. What aspects of Hegel scholarship do you find most exciting or promising right now?*

**RP:** Well, I think some of the things you mentioned are still quite prominent. Aside from the connection with the Marxist and Critical Theory tradition, which made its way from Germany and the Frankfurt School into the United States, one great push toward the so-called Hegel renaissance was Charles Taylor's book in the 1970s. That made clear the relevance of Hegel for what had grown to be a fairly significant dissatisfaction with the liberal democratic tradition, or the deep presuppositions of the liberal democratic tradition, but not because of some Marxist or historical materialist orientation. It made clear the relevance of Hegel's critique of the idea of methodological individualism as a way of doing political philosophy, and the idea that the primary subjects of political thought should be free and rational individuals engaging in contractual relations. Hegel had an enormous influence, and his critique seemed quite powerful to many people, in the 1970s, and that critique is still quite prominent. In recent years it's been transformed into an interest, through Hegel, in so-called recognitional or identity politics, in which merely conceiving of human freedom as the absence of external constraints to pursue one's own interests is a deeply insufficient account of the nature of political life, and the various forms of dependence in human life that are not taken account of by the classic liberal democratic model of free rational choosers ought to paid significant attention.

And even in the sixties there was a kind of neo-Wittgensteinian element to the new interest in Hegel, through the work of Findler, a British philosopher who taught for many years at Boston University. He wrote a book on Hegel that made clear something like the inadequacy of traditional forms of conceptual analysis, and thought that a deeper, richer account of what it is to possess a concept or master a concept was necessary. People informed by the late Wittgenstein and the idea of language games began to see similarities in Hegel, and that's also been a continuing interest. The whole theory of the so-called mind-world relationship, the critique of dualism, the animus against Cartesianism and the priority of the inner mental life of the subject as the first touchstone for philosophical analysis, a whole set of issues that are common to the late Wittgenstein tradition, especially the anti-Cartesianism, especially a philosophy of mind that is not dualistic but not reductionist, not naturalist in the conventional sense that we use the term today: all that began to attract attention.

And then finally, in the mid-sixties, analytic philosophy turned, in an interesting philosophical rather than really historical or scholarly way, to Kant—especially in a very influential book by Strawson called *The Bounds of Sense*, a very quirky but extremely intelligent book by Jonathan Bennett called *Kant's Analytic*, and *The Science of Metaphysics* by Sellars earlier in the sixties. These three books made clear that one could actually do philosophy by thinking about Kant, that it wasn't just a matter of trying to get straight what Kant actually believed, but that there was enormous value for philosophy in trying to reconstruct in a contemporary idiom what Kant was trying to prove. Sellars was one the first to say, "If Kant is here, can Hegel be far behind?" Actually, the subtitle for Sellars' *Empiricism and the* *of Mind* is “Hegelian meditations.” Sellars’ critique of immediacy in *The Myth of the Given* should have, at that point, provoked a great deal more interest in Hegel than it did at the time. Kant is very difficult to read, but Hegel basically invented his own philosophical language, and it takes an enormous amount of work and, to some extent, dedication, as you have to bet that you’re going to get something out of learning how to speak early nineteenth-century German romantic language. So there was a kind of barrier to what would look like the next logical step after the Sellars-Strawson-Bennett phenomenon made Kant a live issue for contemporary Anglophone philosophy. But Sellars kept teaching at Pittsburgh, and every other year he would teach a graduate seminar on Hegel’s *Phenomenology*. So the tradition at Pittsburgh, especially, was kept alive. Recently, two Pittsburg philosophers, Bob Brandon and John McDowell, have begun to write quite a bit about Hegel. Brandon will publish, probably next year, a very ambitious commentary on the entire *Phenomenology of Spirit*. That will be a major book. And McDowell, in some introductory remarks to his famous book of the mid-nineties, said that the entire book was a preparation or a prolegomenon for the study of Hegel’s *Phenomenology of Spirit*. He said that an audible gasp went up in the audience at Oxford when he said that.

So from my point of view, better and more philosophically interesting writing on Hegel began to appear thanks to some extent to the Pittsburg people. This may be much more than you asked for in your question, but it’s quite complicated; there are various dimensions that led people to take seriously again a figure who, except for in the Frankfurt School and the Marxist traditions, was also pretty much forgotten in Germany until the late sixties, when there was another Hegel revival that’s related to this one. There was a group of people at Heidelberg, the so-called Heidelberg School, who began to teach Hegel again as a living philosopher, as a live option. Many American students who were interested in German philosophy went over to Germany to study, but at places like Heidelberg they encountered people who were also very familiar with modern analytic philosophy and were not at all hostile or foreign to the kind of language that made no assumptions about commitments to camps. A lot of students began to come out of the German tradition who had gone there to study.

**YPR:** *You've argued that Hegel is of immense importance to modern philosophy, particularly insofar as he introduces a very different way of thinking into it. How would you characterize Hegel's relationship with that tradition? Is it primarily a critical one? A continuation?*

**RP:** There are several ways to answer that. The most radical aspect of Hegel's philosophy is that not only does he think that the history of philosophy can be done in a philosophical way, but he thinks philosophy is essentially the recovery of it's history in a certain way. So, for Hegel, the wrong thing to think about philosophy is that it is a series of dead ends, which provoke a series of new beginnings, and which, sometime after the 17th century and the invention of modern natural science, finally got on the right track as a kind of formal analysis of the conditions for empirical knowledge and became a quasi-science. Hegel thinks of all fundamental positions in the history of philosophy as partial versions of the truth rather than failed attempts at the whole truth. So that makes the history of philosophy for him significantly more important than for any philosopher in that history, since the proper recovery of the partiality of these other positions is an essential component of what he's doing.

He's the most synthetic, systematic unifier of various positions in the history of philosophy, or at least attempted to be more so, than there ever was or ever has been afterward. The only comparison is someone like Heidegger, who thinks there is a way of narrating the entire history of philosophy as just one profound, colossal failure. Things went very bad right away for Heidegger, and there is an interconnected story to be told about how this early failure overshadowed everything that happened afterwards until a final collapse, at the end of philosophy, in nihilism. Hegel is something like the polar opposite of that.

Secondly, something that people say about Hegel that is a truism but ought to be mentioned is that he tried very hard to reanimate modern, post-Newtonian philosophy with themes from ancient philosophy that he thought were not only compatible but essential to the overall truth of the positions one-sidedly presented by the modern philosophy of Descartes, Leibniz, Spinoza, Hume and Locke. Like many Germans of this period, he had an intense fascination, almost a kind of pathological absorption, in the lost grandeur and beauty of Greek civilization.

Greek philosophy was extremely important to these people, especially when they were very young, and it stayed with them throughout their lives. Hegel sometimes gets carried away by this: in his *Philosophy of Subjective Spirit* he ends by saying that no philosophy of mind after Aristotle is worth anything compared to Aristotle. He shouldn't say that! He should say that the best stuff comes at the end, after it's all developed. He gets so carried away with how fantastic Aristotle is, how Aristotle had avoided all these Cartesian dualisms and spiritualisms and immaterialisms that he says we ought to be thinking mostly about Aristotle. Systematically speaking, he got carried away. So that's the second important aspect of his relationship to the history of philosophy, this profound respect for the vitality of ancient philosophy, not just as a historical residue, but as a vital element that needed to be recovered and revived for the sake of modern philosophy.

And then third, at least as I have argued, is his relation to Kant: that Kant marks the decisive event, misunderstood by Kant himself but nevertheless there in Kant, of a profound appreciation of the novelty of modern philosophy and, as Hegel thought, its potential completion in Hume. There are a lot of ways to characterize this, but the thing that's been the most controversial about this interpretation of Hegel is the view that Hegel is continuing the demonstration that Kant began in showing that philosophy is not about the furniture of the universe, not a way of knowing a priori with pure unaided human reason what the necessary constituents of reality must be. This is an extremely controversial interpretation of Hegel because it seems on the surface that Hegel thinks he can show that the nature of everything underlying the appearances is really cosmic mind or divine mind or something like that. But what I've tried to show is that that's a very superficial reading of those passages, and that if you read what Hegel says about Kant, and the end of this "bad" metaphysics—the substantialization of reason's products as real elements of the world—you will see that the project he's actually about is a continuation of Kant's account, not in the strict Kantian sense of the necessary conditions of the possibility of experience, but in terms of the subjective character of experience. Hegel thinks he can show that experience shouldn't be understood as positive impositions on an alien, exogenously given matter, and that, once we understand the proper relationship between, let us say, "the mind" and "the world," the traditional problems of idealisms on the one hand and realisms on the other fall away. So characterizing Hegel as a metaphysical realist is just as misguided as characterizing him [unproblematically as an idealist]. So I think I would count this appropriation of Kant, this post-Kantian completion of the Kantian destruction of the old metaphysics for the sake of a new speculative philosophy essentially about the nature of human reason and its place in the world, as a decisive element of his appropriation of the history of philosophy. But that's gotten me into a lot of trouble, and I've spent twenty years trying to defend that view.

**YPR:** *You've been doing a wonderful job so far of predicting our questions, since the next one was going to be about Hegel's relation to Kant, and the following one about your "non-metaphysical" interpretation of Hegel.*

**RP:** Well, "non-metaphysical," I should say something about that, too, because that's caused a lot of problems. There was a German scholar named Klaus Hartmann who, in the sixties, proposed what he explicitly called a non-metaphysical view of Hegel. He was referring to something like a category theory, a theory of the basic forms of thought, and I don't think I would characterize my view like that. To say that Hegel's theory is not about the world doesn't mean, in other words, that it's really about human thinking. That's exactly the mistake he's trying to free us from: "Well, if it's not about the world, it must be just about how we think about the world, so that's what you must mean." That's not what I mean! Hegel is, in the grandest sense of all, a metaphysical thinker. He's talking about what the human mind can know non-empirically about the truth. That's metaphysics! To a certain extent, though, he thinks (like Kant) that what the human mind knows without the benefit of empirical experience is itself. It knows the true nature of itself, and it therewith knows its place in the universe. It knows what it is to be a rational being, and why being a rational being is being a part of the whole, not an outside observer of the whole. So I consider all those metaphysical themes, but it was necessary [not to emphasize metaphysics] given the conventional interpretation of Hegel, which is *traditionally* metaphysical, that Hegel is philosophy's attempt to deduce *a priori* the structure of the universe—how many planets there are, all that kind of hogwash. It's metaphysics without the idea that it's telling us what *must* exist and what *cannot* exist in the old traditional sense of metaphysics since Parmenides, since the beginning of philosophy: "If not-being cannot *be*, then, no matter what it looks like, there's no change, and there's no difference between things, because that would be something not being something else, and we know not-being cannot be." So, that kind of metaphysics, the *a priori* deduction of the structure of reality, I think Hegel regards, not as a mistake, but as a partial and incomplete statement of the sufficiency of human reason unto itself in its determination of what it is as reason. So, I'm happy enough to call Hegel a metaphysician, I just don't think he's a metaphysician in the way, say, Leibniz or Plato were metaphysicians.

**YPR:** *Moving away from Hegel now, there's always much discussion of a division between "Continental" and "Analytic" philosophy. What do you think of this conception? Is it meaningful, and if so, what do you see as the concrete differences, and if not, why do people talk about it so much? If there is any difference, where do you see the relationship heading?*

**RP:** I can say, for myself, personally, I've never paid any attention to that, on one side or the other. We were going through my intellectual biography earlier, and the importance of both people like Hegel and Heidegger and people like Wilfred Sellars. I think that if you're exposed in the proper way to people with something philosophically interesting to say, and not given ideological blinders, you just find great philosophy wherever you can find it, and you take advantage of it. I would never for a minute think of not paying attention to Quine or Wittgenstein or Russell or Moore, just as I wouldn't think of ignoring Schopenhauer or Nietzsche or Fichte, and I think that's the attitude anybody ought to have. There are two difficulties, before we get to the big question of "analytic" and "continental." The first is the fact that philosophy has become so professionalized. So sometimes the discussion about analytic and continental philosophy isn't really about philosophy, it's about the way philosophical study is organized in modern American and European universities. This mode of organization is the mode of the natural sciences, and this kind of organization was developed in the nineteenth century, and is a matter of sub-fields and disciplinary specialty. The graduate programs are supposed to ensure that you can be credentialed as an expert in the latest research in the specialty. It's like the model of science: you try to find your way to the cutting-edge where people are doing new things that haven't been done, and you contribute to that by publishing your own little part of it. It's a very interesting topic, and when people talk about so-called analytic and so-called continental philosophy, it's really a question more about what the *professional* study of philosophy has meant. Because the problems that exist in the United States—hyper-professionalism, over-specialization, what the Germans call *Fachidioten*, "discipline idiots," they just know their own sub-specialty, the last five articles published on reference theory or something like that—well, the same thing occurs in Europe, where universities are organized in a more feudal way, not so much on the model of the sciences, but on the model of a great figure who has a school. That's just as deadly, because you're assigned a task in the "great" figure's view of the universe. You know, "You, work on Marsilius of Padua and my problems." So, a lot of this is, I think, a feature of the way graduate culture is organized, and in my opinion, it's organized horribly. And the ethos and culture that are created—they're not the fault of any pig-headed or prejudiced person. It doesn't work that way, although there are plenty of those around on both sides of the Atlantic. But the deeper problem is that philosophical education has become so complicated and specialized that the ability of people just to see good philosophy wherever it is and make use of it has been degraded to some extent by virtue of this. And I think if you fight against it, you have to actually learn more than you've been taught, because you have to get out of the disciplinary boundaries created by professions.

That said, a lot of people raise this topic because of the enormous impact Richard Rorty had on (one could say) "American philosophy," but American philosophy is like "American education:" there are 2,500 four-year colleges in the United States. When you say something like, "Rorty's impact on American philosophy," in terms of his impact on upper-tier elite research university philosophy departments, it isn't very much. I have enormous respect for his work; I think it is, in itself, apart from these other questions about categories, just extremely interesting and deep. But because I work in the European tradition, people are always asking me, "Well, isn't it wonderful now? Everything's changed, everybody's talking to everybody." No, they're not! Nothing's changed; this is all a kind of PR. And so people say, "Well, look at Richard Rorty." Yeah, well Richard Rorty was at a Comp. Lit. department at Stanford; there's nothing different. There are a number of really interesting people around who don't pay attention to these disciplinary bounds: Stanley Cavell until he died, Bernard Williams, or, until he died recently, Richard Wollheim. But again, because of the way philosophy education works—how journals are put out, who reads them and what their orientation is and so forth—there hasn't been much.

Now, substantively for the issue itself, increasingly in Europe the old mode of organizing things around a "master" and a set of texts is rapidly disappearing. So, again, that's a structural issue more than it's a philosophical issue. You have to get two PhDs in most European countries to be a university professor: you have to get a PhD, which many people get and is not necessarily an academic degree, after you get out of college. You write a dissertation, and you get a degree. They don't have graduate school the way we do, where you have a series of courses that you have to take; you show up with a project, you figure out what you have to learn, you go learn it, and you write the project. But then you have to do a *Habilitationsschrift* on a completely different topic. This means, if you want to do a *Habilitationsschrift* on Hegel after you've done one on, say, Leibniz, then you have to learn Greek (because you have to know Aristotle), you have to know Latin (because Hegel knew Latin and he knew the medieval tradition), you have to know all the French and American literature, and then you have to master the vast amounts of literature written on Kant, Fichte, Schelling, and Hegel since they died. Well, that means you're about forty-four years old, literally, when you're finished doing your research. And the Europeans have sort of awakened to the fact that their (basically) thirteenth-century university system isn't functioning all that well. So, to some extent, the whole issue of Continental and Analytic is an anachronism, because Continental philosophy on the Continent is evaporating. The whole Heidegger movement in France and Italy is on its last legs. It is the last gasp of the whole tradition of philosophy as a kind of avant-garde cultural enterprise connected with literature and poetry and art and architecture, which is some of what people mean by "the Continental tradition:" that it's not just a disciplinary matter, but it's taken to be connected to the entire world of culture in a way that, in America, it tends not to be. That may have been true, but that's also evaporating. Thus also the tradition of Hermeneutics that comes out of Gadamer—though it's still quite strong in Italy and some places in Germany—is evaporating. And the pressure on modernizing the university is also changing the way that graduate education in philosophy is organized, and, to some extent, Americanizing it. They're very weird about this: there's a very good German philosopher who writes on Heidegger but is mostly an analytic philosopher called Ernst Tugendhat, and about a year ago he wrote a much-noted article in Germany asking the Germans to ponder the fact that, for the first time in Western history since Leibniz, there is no German philosopher alive to whom everyone in the world thinks they have to pay attention. That's a staggering thought! It is true. Since Heidegger, maybe since Habermas, who is now near retirement, there is no major German figure. They feel terrible about it! They've always had one, and now they don't.

So whatever this distinction was, firstly, I think it's more accurately understood as modes of organizing knowledge rather than commitment issues in ideology or in philosophy. And, secondly, it's changing so rapidly under the influence of forces that are not themselves purely philosophical that it's become a kind of anachronism to talk that way, except with respect to a large group of American philosophers who still identify themselves with this other conception of philosophy as an avant-garde cultural humanistic enterprise.

**YPR:** *This seems to suggest that you have extra-paradigmatic values that allow you to determine what constitutes good philosophy, so that you can move from one paradigm to another to seek out good philosophy. What are those values?*

**RP:** Well, I should say as a beginning footnote that I am in an extremely privileged position because I teach at the Committee on Social Thought at the University of Chicago. I literally get to teach absolutely anything I want, in any discipline. And Chicago has a kind of *ethos* where it's not regarded as "poaching" or as the incompetent trying to parade as experts. Everyone understands the Committee as the attempt to do something different, to break the hold of the traditional way. And if a little amateurism creeps in, or dilettantism, or café society philosophy, well, that's the price for having the Committee on Social Thought and the people at Chicago are willing to pay it. But what I mean is that it's easy for me to say this, because I have an established career, and I went through a long career of just doing traditional, disciplinary philosophy. So, institutionally I'm in a very unusual position: even people who have interests similar to mine, if they're in a philosophy department, they have an undergraduate program that has to be staffed, certain courses that have to be filled every year, and somebody has to teach Intro to Ancient, somebody has to teach nineteenth century, somebody has to teach Kant every other year in the graduate program, and if you're that person, you have to do that. Your research is just determined by the size of the department and what you have to do. So it doesn't matter, in a way—going back to this institutional question, which I think is a neglected one in discussions of philosophy—what your new paradigm for philosophy is, because you don't get to do it. And if you're doing it before you're tenured, then you're taking an even huger risk that you won't be regarded as having done real philosophy. And the same thing is true with your promotions: your status in the profession is what gets you offers from other places, which boosts your stock, which increases your salary, and so the incentives built in to staying within disciplinary boundaries until you get to a certain point—when you can just raise the Jolly Roger and come out of the closet or whatever—are so huge that this whole practical problem ties in.

That said, I think the most interesting questions in philosophy, perhaps the only real, significant questions in philosophy, are all normative. They're "ought"-questions to a certain extent, and they're basically about two things: what ought we to believe, and what ought we to do. And I don't think texts simply written in philosophical prose are the only form of human reflection about that topic. So I think a lot goes on in Shakespeare, in Henry James, in Edouard Manet, for example, that ought to be considered something other than mere examples of philosophical ideas. It would take a long time to defend this, or make it even credible, but at the end of the day, what I think requires a form of reflection that's not limited to the disciplinary prose of philosophy is this set of normative questions. What is it to lead a free life? How do you find that out if you're a philosopher? "Well, you examine the concept of freedom." And where does that come from? Which concept? The Aristotelian one, the Scholastic one, the Lockean one, the Hegelian one, the Rousseauian one? If they're all competing, what are they all about? I don't think reflection on that topic without, say, the nineteenth-century novel makes much sense. People try to figure out whether their life is their own, and if it isn't, how to recover it. How can you understand that topic without thinking of what these geniuses, who try to portray, let's say phenomenologically, the struggle to come to understand the question of whether their lives are their own? They were smart guys! They weren't just excellent craftsmen, they were profound thinkers. So, for me, the principle of looking for interesting philosophy wherever you can find it means not only looking across this so-called Analytic-Continental divide, but *wherever* you can find it. I mean, you have to do some work, and you have to learn to think about it in a certain way, but that's possible. You don't have to be stuck with, you know, "S à P" all the time. It's an important way to think, but you don't have to always *keep* doing that.

**YPR:** *Karsten Harries once said something to this effect in a lecture: "In a sense, modern art is more philosophy than art, and modern artists are often more like philosophers than like what we traditionally understand as artists." Given what you've just said, how would you characterize the relationship between philosophy and art?*

**RP:** Well, Karsten's the one you should ask. He has a very beautiful book on the subject called *The Meaning of Modern Art*. It's a wonderful book, and a lot of the answer is in there.

But my favorite art historian is Michael Fried at Johns Hopkins, who I consider as much of a philosopher as anyone else. And one of the things that Fried is interested in, among many, is how visual intelligibility is portrayed in painting: what an embodied mode of the vision and the intelligible presence of the world to a mind is like. And he knows how to bring that issue to life in the analysis of painting. It takes great intellect and great sensitivity to do it. There are a lot of issues like that, which require philosophical reflection if the art object is to be properly understood. Something happened to the history of Western art around the middle of the nineteenth century, and not just painting: literature, poetry, architecture, and music all began to explore issues we would now label as much more "cognitive" issues, much more about their own possibilities than the traditional canon of aesthetic values, which was primarily about beauty—understood either as some version of the pleasant, or some version of an attempt at perfection. But something happened that made those categories inadequate to art. There was an implicit claim presented by, let's say, painting starting with Manet, that the world ought now to be looked at like *this*, not like *that*, not like the way it had been looked at. I went to a MoMa show where they exhibited *The Execution of Maximilian*, which is an excellent painting by Manet. And it looks like a normal realist painting, except you'll notice that the man who raises his sword as he's about to give the signal to shoot is behind the shooters, and he has red pants on. And between the pants of the executioners, you should see his red pants. But all Manet does is take his brush and just put a tiny little red mark there, completely unintelligibly in terms of classical realist notions of painting. And where Maximilian's hands are gripping the two generals he's being executed with, instead of painting his hand—all the other ones are painted—Manet just puts a glob of flesh-colored paint there. And I don't think traditional categories of aesthetics are adequate here. Philosophers are always saying, "This attention to art and aesthetics is ridiculous. Art doesn't make any claims. There are no propositions, there are no arguments." Well, what's the claim made by not painting that leg? By just putting a bit of red paint there, as if to indicate, "I could paint it if I wanted to, I want you to know that, but I'm not going to." What's the point of that? If you're interested in what the point of that is, then you've got to be interested in why somebody thinks it's valuable to do it that way, and that's got to make you think about what's valuable about painting, and that's got to make you think about value. So I don't think it's possible to understand painters' presentations of how the world *ought* to be experienced, if it's to be experienced in its truth, without philosophy. Art criticism and art history that don't have that dimension seem sterile to me, mere academicism. It's just a listing of influences, iconography, painting technique, how fast he could paint, the nature of the brush stroke: that's all very important, but I think it's certainly possible to raise what we in philosophy would call propositional claims and certain considerations relevant to accepting or rejecting them.

It's the same sort of thing with literature. I tried to make clear in a book on Henry James how much philosophy is impoverished if it does not pay attention first to the issue of the proper act-description of the moral context in which a philosophical problem is posed. What exactly is happening? [James'] characters' self-reports about what they intend to do and what they think is happening are tremendously untrustworthy, and if philosophy doesn't take that into account somehow and try to find a way into, say, the density of moral complexity in Henry James—density that comes from the difficulty of getting a description of the character's "mindedness" properly set—you're not going to have very interesting philosophy. You know, trolley train examples are all perfectly appropriate, they can be enormously powerful, but if you really want an example to get your wheels spinning, try *The Ambassadors* or *Remembrance of Things Past* or something like that. Again, not merely as a propaedeutic, but as a way of thinking about philosophical considerations relevant to claims of truth in a different way.

**YPR:** *This might be a terrible question. How can Hegel be an aid to us in dealing with the so-called "problems of modernity?" Is there something in Hegel's relationship to modern philosophy that speaks directly to some of these problems?*

**RP:** I think the main theme in Hegel's social philosophy is the proper way to understand human dependency—that the ideal of being as self-sufficient and having as much self-direction of your life as possible is a false ideal, a kind of pathology. Hegel has tried to provide a different sort of framework for understanding, let us say, the relationship between dependence and independence of human subjects considered as agents. He's aware of all the problems that that generates, of raising this point that there are aspects of human dependency that are pre-volitional. You didn't contract to be in the family you're in, and nobody asked you to be in it. But there's something wrong with taking that as justification for acting as if you weren't in one. There's something wrong with it, and it's hard in the traditional language of liberal-democratic politics to say what that is. "If you don't want to be in it, then just opt out, don't be in it anymore!"

Now, Hegel's suspicions are very deep about that kind of reaction. But that immediately raises the counter-suspicion that he's an anti-individualist, an organicist, a holist. And after the Second World War, especially in Anglophone literature, there were a whole series of political philosophers and theorists looking for the "dark side" of the German soul. "How could this happen in the most civilized nation on the planet, the most advanced in high culture, music, art, literature?" People began looking in their philosophical tradition, and noticed something that is, in a very general sense, true: that the advent of modernity and modernization in Germany was very different than it was in France or England, that it was, in a way, resisted more, that the German Enlightenment was much more friendly, from the Berlin years on, to religion than the French or English versions, and that there was just considerably more anxiety and skepticism, sometimes traced to the fact that industrialization in the latter half of the nineteenth century in Germany was much more rapid than it had been over the 150 years prior in France and England. In Germany it just happened in a kind of shocking, traumatic way. So these factors came into play as people were looking for indications of a German resistance to real modernity, to individualist, free-market, liberal-democratic modernity, and Hegel was chosen (by Karl Popper, Sydney Hook, and quite a lot of other writers in the fifties) as the benchmark, the clearest sign, of the excesses of German Romantic political philosophy. Even Isaiah Berlin, an enormously influential postwar intellectual in the English world, had his heroes in people like Herder, but he had his enemies too, and Hegel was one of them. A whole bunch of people saw Hegel as the indication of where Germany had gone wrong.

So Hegel's social philosophy existed under this cloud, and it still does. Any sensible person realizes that political life is decisively influenced by passions and emotional factors that are pre-deliberative and pre-voluntary. You don't get to just decide you ought to have a certain emotion (perhaps loyalty or solidarity), to think you should have it, come to see the argument for why you should have it, and have it. We all know it doesn't work that way. But, very often in liberal-democratic culture, any conclusion drawn from that looks like the beginning of a slippery slope towards paternalism, manipulation, propaganda—and we're terrified of that, especially in the United States. This aspect of how to think about the political passions is the subject of the Castle Lectures I'm giving next year: I'm going to be talking about this problem inflected in great, classic American Westerns.

So, it has been hard for the Hegelian alternative to be debated openly, because the suspicions about the direction this is heading are so intense. The same thing happened with the phenomenon of the Communitarian critique of liberalism in the eighties and early nineties that enlisted Hegel as an ally. That seemed to many people just to be a retreat from reason, a way of validating forms of association and belongingness that are irrational—that may exist, but that always have to be open to rational reflection and deliberation. And the Communitarians seemed to think "No, that's an exaggeration, you can't do that." Hegel was roped into that, so it looked like another indication that, by being against classical liberal theory, Hegel was *illiberal*.

So, one of the things I'm interested in doing is raising the kind of question that Hegel wants to raise in a way that doesn't immediately provoke this knee-jerk reaction. My former colleague at Chicago, Stephen Holmes, a political theorist, is very typical of this: any deviation in the slightest from liberal-democratic orthodoxy is a slide towards the Dark Side—it will get you if you stray one foot from the path. It's a fair suspicion—the cloud is there in some respects—so it's up to those of us who think that there's something in Hegel that might be of value that isn't illiberal to articulate it in a way that doesn't immediately ring all these bells.

**YPR:** *Now, for something a little more fun: if you were able to ask one figure in the history of philosophy a question, what would it be?*

**RP:** Well, I would ask what Aristotle means by the "active intellect." I would ask Plato whether there are ideas of ideas. I would ask Kant what he meant by the B160 footnote—this is where Kant says, "You know this distinction I make between intuition and concept, as if they're these two radically different things? Well, you know, they really aren't. The forms of intuitional unity are actually also unified by the work of apperceptive spontaneity. I had to start somewhere, so I started with intuitions, but you shouldn't think of them as distinct." That sort of seems to undermine the entire strategy of the deduction of the structure of critical philosophy, and I would ask him what he meant by that. I would ask Hegel what he meant when he said, in the introduction to the *Phenomenology*, that "beauty hates the understanding for asking of it something it cannot do," a beautiful thing to say. What did he mean when he said, "The wounds of Spirit are self-inflicted but can be healed, and in fact, can be healed without scars?" I could go on.

What I worked on for a couple years was, what would Hegel have said about modernist, especially abstract, art? Because Hegel has this famous thesis—which, on the one hand, is just banally obvious—that, compared to Romanticism and the Middle Ages and Greek culture, art in the bourgeois world is no longer the vehicle for the highest expression of human self-knowledge. So, he says something like, "Art is therefore for us a thing of the past." Now, he can't have meant that people would not want to make art. And there's something about what he says that history seems to have proven right: the Saatchi gallery in London, sharks in tanks, the Danto phenomenon, it just looks like the endless trivialization of high art. We've reached a cycle where even the art market can't keep going forever. But, given that he didn't mean *that*, and given that he doesn't mean "surpass" in the sense of "go beyond and leave behind"—he never meant that by *Aufhebung*, a technical term for surpassing while including and bringing forward—what did he really think about art in the bourgeois age, the age of the non-heroic, the age of small businessmen and bureaucrats and teachers? What did he think the function of art was? Granted that it's not the vehicle of the highest feelings, what's second-highest? I would like to know that.

**YPR:** *Finally, can you say a little bit about your current interests in film?*

**RP:** Well, I'm working on this project that I'm very excited about, about Westerns and the American self-understanding. Westerns are mythological and epic narratives. They're not terribly modern: the characters in them just are the characters they are. Good guys don't struggle with having to do bad things, and bad guys with the occasion to [do good]. They're types, and they're presented in a pathological way. So there's a way in which the very deepest, basic issues of American self-understanding, in the hands of people who are real geniuses (like John Ford, Howard Hawkes, Anthony Mann), manifest themselves for us. What we think of ourselves—especially what we think of the American founding, the push to the frontier, the conquering of the land from Native Americans, and the nature of the allegiance that civilized order creates when the crucial effect of the possibility of civilized order is the monopoly on the use of legitimate violence. So I think there are ways of getting at that question though film.

Support of this publication was assisted by a grant from the John K. Castle fund. The Castle Lecture and Publication fund was endowed by Mr. John K. Castle. They honor his ancestor the Reverend James Pierpont, one of Yale's original founders.

Special thanks to the Yale Philosophy Department and the UOFC.

This issue of the YALE PHILOSOPHY REVIEW was produced using Adobe InDesign <http://www.adobe.com>, and was printed by the Yale Reprographics and Imaging Services of New Haven, Connecticut. The typface is Times.

YPR

2007