

Meditations

The Undergraduate Journal of Philosophy at UCLA

Issue 5, Spring 2017

Meditations

σύννοια

The Undergraduate Journal of Philosophy at UCLA

Issue 5, Spring 2017

Editor in Chief
Shushan Ginosyan

Selection Committee

Rian Bahadurji

David Dixon

Andranik Khacheryan

Pablo Mendia

Max Schoenberger

Craig Wierman

Editorial Board

David Dixon

Andranik Khacheryan

Max Schoenberger

Craig Wierman

Eva Yguicu

Department Staff Assistance

Blake Livesay

Tanya Kim

Faculty Advisor

Andrew Hsu

Graduate Advisor

Michael Hansen

Editor Emeritus

Evan Green

Contributing Authors

Blake Deal, UCLA

Blair Kimble, UCLA

Shayan Koeksal, SFSU

The publication of Meditations is made possible by the financial support of the Undergraduate Philosophy Club at UCLA and the UCLA Philosophy Department.

Letter from the Editor

Dear Reader,

Serving as Editor in Chief of this year's journal has been an extremely gratifying experience for me. As an admirer of the journal since its inception, I sought to sustain the vision of the journal that had been established by its founders, while also adding a few new touches of my own.

This vision of the journal could not have come to fruition without the support and dedication of the Editorial Board and Selection Committee. These individuals spent countless hours reading various submissions, debating their merits, and once the selections had been made, working with our authors to ensure that our journal published the best possible papers. I have no doubt that this journal would not exist if it were not for the dedication of these individuals, to whom, I and all the beneficiaries of the journal owe a great deal of gratitude to. I'd also like to thank our authors for submitting their truly unique works to our journal, and working assiduously with the Editorial Board to prepare their papers for publication. Your perseverance and creativity is truly inspirational, and I hope your experience with the journal has been one of value.

This letter would not be complete without a warm thanks to Austin Beltrand, the President of the Philosophy Club at UCLA, for his guidance and support throughout this entire process. His continued support fortifies the strong relationship that exists between the Meditations Journal and the Undergraduate Philosophy Club at UCLA.

Our pool of submissions this year preceded those we've received in years past—both in quantity and quality. This is a telling sign that the Meditations Undergraduate Journal of Philosophy awaits a bright future. I have no doubt that my predecessor's passion and diligence will catapult the journal to greater heights. With that said, I sincerely hope you enjoy reading the 5th edition of the Meditations Journal!

Respectfully,

Shushan Ginosyan

Editor in Chief, Meditations

Table of Contents

Letter from the Editor Shushan Ginosyan	4
Did a Picture Hold Rorty Captive? Blake Deal	7
Promising and Its Necessary Relational Component Blair Kimble	26
The ‘Now What?’ Problem for Moral Error Theory: Substitutionism vs. Revisionary Moral Expressivism Shayan Koeksal	38

Did a Picture Hold Rorty Captive?

by Blake Deal

“A picture held us captive...” (*PI* §115), so says Ludwig Wittgenstein in his book, *Philosophical Investigations*. This picture includes a number of concepts and methodologies entrenched in the western philosophical tradition since the 17th century, ranging from theories of mind and epistemology, to language and logic. Wittgenstein’s little phrase summarizes the dissatisfaction many contemporary philosophers continue to express with traditional philosophy and their aim to transcend the picture to which their predecessors devoted themselves for centuries.

Few philosophers have done more to expound upon and undermine this picture than Richard Rorty. According to Rorty in his book *Philosophy and the Mirror of Nature (PMN)*, “The picture which holds traditional philosophy captive is that of the mind as a great mirror, containing various representations—some accurate, some not—and capable of being studied by pure, nonempirical methods” (*PMN* 12). In a word, the picture which held philosophers captive since around the time of Rene Descartes¹ is *representationalism*. Representationalism is the idea that the mind interacts with the outside world by means of inner representations. Epistemology consists in devising nonempirical methods to discover which of our representations accurately “mirror” reality, and therefore which of our representations can justify our beliefs. If philosophers can discover how to differentiate privileged, accurate representations from inaccurate ones, then philosophy can provide a “tribunal of pure reason” (*PMN* 4) able to adjudicate claims to knowledge (*PMN* 3) made by the prevailing culture.

In *PMN*, Rorty intends to escape this picture:

“The aim of the book is to undermine the reader's confidence in ‘the mind’ as something about which one should have a ‘philosophical’ view, in ‘knowledge’ as something about which there ought to be a

¹ Richard Rorty, and, as we shall see, Hubert Dreyfus and Charles Taylor, portray Descartes as a representationalist, but this is still

‘theory’ and which has ‘foundations,’ and in ‘philosophy’ as it has been conceived since Kant.”

(*PMN* 7)

Hence, Rorty’s project is primarily deconstructive. Rorty does not intend to refute any particular theory and replace it with his own theory, but rather to demonstrate that these concepts of “mind” and “philosophy” are products of historical contingencies. None of the problems of “philosophy” are perennial. Contrary to Immanuel Kant, human reason *does not* have the peculiar fate of being “burdened with questions which it cannot dismiss” due to “the nature of reason itself” (*CPR* 99). Rather, how we conceptualize things like “the nature of reason itself” or “mind” are themselves products of a unique historical development, not a product of the nature of reality. One step towards dissolving the representationalist picture is rejecting the legitimacy of searching for these “nonhistorical conditions of any possible historical development” (*PMN* 9). As Rorty argues throughout *PMN*, the puzzles our intellectual tradition now burdens us with are optional and could have developed in a number of other ways.

However, although Rorty purports to escape the representationalist picture, Hubert Dreyfus and Charles Taylor argue that Rorty is still entrenched in it. In their book *Retrieving Realism (RR)*, Dreyfus-Taylor attempt to undermine representationalism like Rorty, but do so by articulating a “contact theory” of realism. Their contact theory consists in reshaping concepts leftover from the representationalist picture to accommodate the insights of philosophers like Martin Heidegger and Maurice Merleau-Ponty. Dreyfus-Taylor make a compelling case that Rorty is still wrapped up in the very picture he claims to transcend. For example, Dreyfus-Taylor characterize representationalism as having an inside-outside structure (I/O). The puzzle representationalism presents us with is how our *inner* representations can represent the *outside* world. Likewise, representationalism has a mediational structure. We can only know the outside world *through* or *by means of* our representations; thus, we are not in direct contact with reality. Dreyfus-Taylor claim that because Rorty retains this I/O distinction, and retains mediational, nonrealist tendencies, he is still mired in the very representationalism he rejects. But ironically, Rorty later accuses Charles Taylor of retaining

aspects of representationalism, since Taylor invokes (supposedly) antiquated concepts like “reality.” So who is right?

In this paper I will argue that, although Dreyfus-Taylor are right to point out how Rorty somewhat retains a I/O and mediationalist schema in a number of passages in *PMN* (not to mention his many other writings), Dreyfus-Taylor’s criticism is not as substantial as they think. Dreyfus-Taylor’s accusation against Rorty’s alleged representationalism, even if true, reveals little more than a methodological difference between each group. The fact is that each group shares approximately the same goals and arrive at approximately the same conclusions. The difference between these philosophers stems from the fact that Rorty’s project is primarily *deconstructive*, whereas Dreyfus-Taylor’s project is primarily *constructive*. Dreyfus-Taylor are “maximalists” while Rorty is a “minimalist” (*RR* 41). Rorty does not wish to replace representationalism with anything, he just wants us to forget about these issues altogether, whereas Dreyfus-Taylor replace the representationalist picture with another picture, or at least a different approach. The program each group offers in conclusion are nearly identical, but each group disagrees over what to call it.

If I succeed, I will have demonstrated that the long-lasting disagreement between Rorty and Dreyfus-Taylor is one of labels, methodology, and emphasis, not so much theory and content. To demonstrate this, I will first provide an example, in context, of where Rorty appears to still be mired in the I/O picture, explain the implications Dreyfus-Taylor draw from this, and lastly, compare and contrast key passages in their writings. My discussion will demonstrate the fact that their debate is, by in large, a long-standing—and somewhat confused—verbal dispute.

According to Rorty, the key portion of *PMN* is found in chapter four, where he attempts to refute the validity of “privileged representations” and thus to undermine the analytic project altogether. As Rorty sees it, analytic philosophy was based upon finding (1) privileged representations and (2) privileged meanings, according to which we can justify our claims. For the longest time, philosophers accepted Kant’s distinction between “necessary” or “formal” aspects of our cognition, and “contingent” or “material” aspects of our

cognition. Corresponding to these are “analytic” (necessary) judgments and “synthetic” (contingent) judgments. Analytic judgments are those judgments which are true by virtue of meaning alone, whereas synthetic judgments synthesize experience with concepts. In order to produce knowledge, we must take disorganized empirical content (empirical intuitions) and organize them by subsuming them under concepts. Finding privileged representations consists both in discovering which of our judgments follow correct cognitive procedures, or in discovering which of our judgments are “true by virtue of meaning” (*PMN* 169) alone.

However, Rorty contends that Wilfrid Sellars and Willard Van Orman Quine each undermined one horn of the analytic project. In its place, Sellars and Quine offer *holistic* theories of justification. According to Rorty, Sellars refuted “the Myth of the Given,” the idea that empirical content (empirical intuitions) exist in sharp, disorganized contrast to concepts, while Quine undermined the analytic-synthetic distinction, and along with it, the idea that judgments can be true by virtue of meaning alone. Once Sellars and Quine dissolved these sharp dichotomies, there were no privileged representations left to appeal to. As James Tertaglia explains, Rorty believed he demonstrated “that there is no way to short-circuit the judgement of society by appealing directly to representations in the Mirror of Nature. Just as Sellars showed that sensory givens cannot do the job, so Quine showed that analytic truths [linguistic givens] cannot do the job” (*JT* 124). Hence, there are no meanings or inner representations which afford us special epistemic access to reality. Without such representations, we can only justify our claims by appealing to the customs and practices of our own cultures.

It is at this point where Rorty makes one of his characteristically “representationalist” statements: “For the Quine-Sellars approach to epistemology, to say that truth and knowledge can only be judged by the standards of the inquirers of our own day is not to say that human knowledge is less noble or important, or more ‘cut off from the world,’ than we had thought. It is merely to say that nothing counts as justification unless by reference to what we already accept, and that *there is no way to get*

outside our beliefs and our language so as to find some test other than coherence.” [emphasis mine]
(*PMN* 178)

It is here where we can see Dreyfus-Taylor’s I/O distinction creeping into Rorty’s thinking. Why are the criticisms of Sellars and Quine so important? Because it shows we no longer have any “givens” left to appeal to; all we have left are our own beliefs and conceptions, and those acceptable to our society. On the representationalist picture, in order to “get outside” of ourselves and our own conceptions, we had to devise methods of discovering which of our representations formed according to proper procedures, independently of our valuations. But now that there is no legitimate way of separating meanings from language, or representations from preexisting concepts and valuations, we cannot *rise above*, *escape*, or *get outside* these. All we are left with are beliefs, language, and the significance our culture imputes to them. Hence, again, in the words of Tertaglia, we can no longer purport to “short-circuit the judgement of society” (*JT* 124). Therefore, Rorty concludes that “justification is not a matter of a special relation between ideas (or words) and objects, but of conversation, of social practice” (*PMN* 170).

Dreyfus-Taylor find this comment of Rorty’s, and those similar to it, especially problematic. Since comments like these found in *PMN* exemplify the same I/O structure indicative of the representationalist picture, they demonstrate Rorty’s inability to escape from this picture. In response to this passage, Dreyfus-Taylor contend, “This is the standard picture, out of which nonrealist theories were generated in the first place. And here we find it invoked within an argument which is meant to repudiate that picture. This is what it means to be held captive” (*RR* 58). As Dreyfus-Taylor see it, Rorty’s implicit adherence to representationalism marks the fundamental flaw in his approach, and colors all of his conclusions with a hue of skepticism and nonrealism, as much as Rorty himself repudiates such labels.

According to Dreyfus-Taylor, Rorty’s talk of failing to get “outside” our beliefs and language also entails a strict demarcation between the Sellarian “space of reasons” and the “space of causes,” which is yet another product of representationalism. The space of reasons is the metaphorical space within which

justification, judgments, and beliefs operate, while the space of causes is the metaphorical space within which objects engage in causal interactions with one another. The space of reasons operates “inside” the subject’s mind, whereas the space of causes operates “outside” the mind, in the world, independently of human agents. Because Rorty holds that there are no privileged representations which follow proper procedures according to the structure of the mind (or language), we cannot escape the space of reasons. Since we cannot escape the space of reasons, and since we no longer accept the idea that we have access to privileged representations which bring us in contact with the “outside” world independently of our judgments and beliefs, the only standard of justification we are left with is the judgment of our own culture. We cannot “find some test [for our judgments and beliefs] other than coherence” (*PMN* 178) because whatever we end up appealing to will likewise be constituted by our own judgments and beliefs.

To combat Rorty’s conclusions, Dreyfus-Taylor attempt to dissolve the strict demarcation between “the space of reasons” and “the space of causes,” between “inside” and “outside,” and to place human subjects back squarely in the midst of their environment.² Instead of conceiving the human subject as a disengaged cogito, Dreyfus-Taylor emphasize our capacity for “embedded knowing” (*RR* 58). The paradigm of human knowledge should no longer be that of a disengaged subject forming beliefs about the world which can be developed into explicit theories and propositions, but rather that of an engaged agent skillfully coping with their environment.³ For example, if we ask Johnny to go into the other room to see if a frame on the wall is crooked, Johnny does not begin theorizing about the frame, or “check the...belief that the picture is crooked against his own belief” (*RR* 59). He just goes and checks it. In order to check it, Johnny exercises a

² See also: Taylor, Charles. “Rorty and Philosophy.” *Richard Rorty*, edited by Charles Guignon and David R. Hiley, Cambridge, 2003, 158-180.

³ Dreyfus-Taylor’s conception of “skillful coping” originates from the writings of Martin Heidegger and Maurice Merleau-Ponty. For example, Heidegger draws a distinction in his book *Being and Time* between two ways of knowing, what he calls “present-at-hand” (*Vorhanden*) and “ready-to-hand” (*Zuhanden*). Knowledge that is present-at-hand is explicit, contemplative knowledge. It is the knowledge one engages in when one theorizes about the nature of hammers, what they are made of, and so forth, whereas ready-to-hand knowledge is when one simply uses the hammer. A carpenter who uses a hammer to pound in nails does not contemplate the properties of the equipment she is using, but simply incorporates it into her activity. “Skillful coping” is another way to describe Heidegger’s “ready-to-hand.”

number of capacities like focusing on the frame, seeking a stable background, and so on. Oftentimes, like Johnny, we have developed these capacities to such an extent that “we no longer notice what we are doing” (RR 59). Most of the time we are absorbed in our own activities without any high level of intellectual contemplation. Hence, the world is not strictly divided into higher order beliefs and concepts on the one hand, and subject-independent objects with their causal relations on the other. Much of our experience is informed by preconceptual know-how involving the interplay between ourselves and our surroundings. This leads Dreyfus-Taylor to emphasize, contrary to Rorty: “*Of course* we check our claims against reality” [emphasis mine] (RR 59); we are in direct contact with it.

So now it all becomes clear, think Dreyfus-Taylor. The reason why Rorty argues about not being able to “get outside” (PMN 178) or to “step outside our skins” (JT 120) is because Rorty retains aspects of the representationalist picture. What causes him to say we can never check how our beliefs mesh with reality itself is that he retains a view of “reality” and “truth”⁴ as things which we cannot *directly*, but only *indirectly*, cognize. Since past explanations of how we indirectly cognize “the truth” or “reality” are moot, we should stop talking about them says Rorty. But, since this is only true according to representationalism, Dreyfus-Taylor remind us, we can devise another picture⁵ within which we can speak about “truth” and “reality” meaningfully.

However, the dialogue between Dreyfus-Taylor and Rorty is more complicated than this conclusion allows. There are a number of things I can say on the topic, but I would like to offer just two arguments that will poke holes in Dreyfus-Taylor’s triumphant account of the situation. First of all, I fear each party speaks past one another when discussing the murky concept of “truth” or “reality.” Since Dreyfus-Taylor and Rorty interact with one another from the standpoint of different pictures, what one group means by “truth” is

⁴ I take these terms to be interchangeable.

⁵ Dreyfus-Taylor would probably not call their observations another “picture” as such. Their observations are phenomenological and so cannot be construed as a pre-established, a priori picture within which they shoehorn reality. I use “picture” here to describe their alternate approach to the issues they discuss.

different from what the other means by “truth;” the same goes for knowledge and justification, but for the sake of space I will only focus on truth. Secondly, due to this equivocation, neither Dreyfus-Taylor nor Rorty can see their mutual similarities. As controversial as the claim may be to both parties, I hold that the goals *and* conclusions of each party are virtually identical.

In their second to last chapter of *Realism Retrieved*, Dreyfus-Taylor help prove my hypotheses in two short pages. While discussing scientific progress, Dreyfus-Taylor question whether or not our increasing ability to control phenomena “shows that our theories correspond more and more closely to the structure of the universe” (RR 145). Dreyfus-Taylor argue that it clearly does. But, pragmatists of the Rortyan variety would object to Dreyfus-Taylor, arguing something along the lines of “we cannot ever hope to speak the language of the universe.” Contrary to getting closer and closer to understanding the underlying structure of the universe,

“[F]or pragmatists, control is not evidence for correspondence [with the underlying structure of the universe]....On this view, rather than control providing a sign that we have got the underlying structure right, what we mean when we claim we have got the underlying structure right is that we have gained more control.” (RR 146).

In other words, for the pragmatist, Rorty included, all we have accomplished is gaining more control over phenomena; we have not pierced into the inner “structure” of the universe. In response, Dreyfus-Taylor retort:

“How could one convince them, then? *Perhaps nothing can convince someone if he or she wants to remain in the Kantian inside-outside picture.* But we can challenge whether the distinction nonrealists are relying on here, between what ‘works’ and what brings us closer to the truth, makes sense of the actual practice of scientists.” [emphasis mine] (RR 146)

By Dreyfus-Taylor’s own confession, those who draw a sharp distinction between successful coping and gaining insight into the structure of the universe view the world from a Kantian (representationalist) picture.

Therefore, we must conclude that when Rorty speaks of not being able to discover the inner structure of the universe, he is not speaking of it from Dreyfus-Taylor's picture. This is the equivocation I am talking about. Dreyfus-Taylor present skillful coping as increasingly revealing the structure of the universe according to their revised Heideggerian- inspired perspective, whereas when Rorty speaks about the structure of the universe, he speaks about it from a highly specialized, representationalist perspective, meant to emphasize the shortcomings of representationalism by its own lights.

Although neither Rorty nor Dreyfus-Taylor clarify what exactly they mean by the "correspondence" theory of truth, I imagine it means something like this: The truth is analogous to Kant's noumenal thing-in-itself. A representation or proposition is true if and only if it adequately *corresponds* to the thing-in-itself, i.e. objects as they are in themselves apart from human cognition. Correspondence is a type of relation between our representations and the world. But something that is strange about the debate between Rorty and Dreyfus-Taylor is their assumption that the correspondence theory of truth is or was as pervasive as they imply. Other figures exist within the western philosophical tradition who rejected the correspondence theory of truth, so the idea that its rejection entails a rejection of the entire tradition is not plausible. Kant himself, from the transcendental perspective, rejected the correspondence theory of truth. Within the parameters of empirical realism, we study the relations between objects and other objects as we do in normal science. But within the parameters of transcendental idealism, we examine the a priori aspects of the relations between our representations and objects.⁶ In this way, transcendental idealism follows *horizontal* or *procedural* objectivity, but not *vertical* objectivity⁷. Meaning, we come to conclusions about objects in reference to a set of a priori procedures, but these objects end up being constituted by our own minds. Therefore our representations do not have any relation to the "outside" world but only relate to one another and to the a

⁶ I heard this distinction first taught by Professor Tyler Burge in his course on Immanuel Kant during Spring Quarter 2016 at UCLA.

⁷ Horizontal objectivity refers to the relation between different representations, whereas vertical objectivity refers to the relation between representations and the objects in the world which they represent.

priori rules governing our minds. How our representations correspond to mind-independent objects was not a concern of Kant's. What this shows is that Kant, the quintessential representationalist according to both Rorty and Dreyfus-Taylor, did not engage in a discussion of the variety that Rorty and Dreyfus-Taylor engage in. Kant was not concerned about how our thoughts correspond to the world, but only how our mind constitutes it, which leads me to my next question.

If Kant himself did not subscribe to the idea that we can cognize things in themselves, or rise above culture to "a view from nowhere" (which is one way of characterizing the correspondence theory of truth), then why do Rorty and Dreyfus-Taylor think correspondence is such an integral part of our philosophical tradition? Both groups assume their discussion of truth and correspondence is essential for overcoming a picture which was first mentioned in the philosophy of Descartes and reached its zenith in Kant, but they do not provide any explicit definition of this theory of truth or provide much evidence demonstrating that this theory was widely operative in past philosophical systems. They present ideas of truth which have something to do with having our concepts, ideas, and representations correspond to *things in themselves* quite apart from our own individual and cultural perspectives, and that in order to attain to this ideal we must discover legitimate methods enabling us to gaze upon objects naked and undefiled. But was this the common picture since Descartes? Perhaps. Nicolas Malebranche, one of the most avid proponents of Cartesianism, says something similar to this picture of correspondence in his *The Search After Truth*:

"I think everyone agrees that we do not perceive objects external to us *by themselves*. We see the sun, the stars, and an infinity of objects external to us; and *it is not likely that the soul should leave the body* to stroll about the heavens, as it were, in order to behold all these objects" [emphasis mine] (*SAT* 217).

Just as Donald Davidson says "we can't get outside our skins" (*RR* 58) and Rorty alongside him says "there is no way to get outside our beliefs and our language" (*PMN* 178), Malebranche in the 17th century says "it is not likely that the soul should leave the body" to perceive objects "by themselves" (*SAT* 217). This

quotation exemplifies the idea that our contact with reality is *mediated* by something else, whether it be our own ideas or the ideas of God. Perhaps it also shows that previous philosophers held that, since we are not in direct contact with reality or objects, we must rely upon God's veracity to believe our ideas adequately represent or correspond to the world as it is *in itself*. But, neither Rorty or Dreyfus-Taylor enter into a discussion of how past philosophers viewed truth. Both refer to "the picture" which "held us captive," and make a few comments about Descartes and Kant, but not exactly how truth and reality enter into their picture, or even if many past philosophers subscribed to this picture.

I contend that, whatever one makes of this correspondence theory of truth, both Dreyfus-Taylor and Rorty eagerly reject it anyway. Both thereby "eliminate the Greek contrast between contemplation and action, between representing the world and coping with it" (*PMN* 11). But, whenever *Rorty* mockingly refers to "reality" he is speaking of it in this specialized, theory of correspondence sense, whereas Dreyfus-Taylor speak of "reality" in an everyday sense of skillfully coping with the world. To recap, (1) both parties share the same goal of undermining representationalism; (2) both parties wish to "eliminate the Greek contrast...between representing the world and coping with it" (*PMN* 11); but (3) whereas Rorty wishes to drop the subject of "reality" altogether, Dreyfus-Taylor *replace* the definition of "reality" from having to do with representationalism, to having to do with skillfully coping in the world. According to Dreyfus-Taylor, we are in direct contact with reality when we find better and better ways coping with it, whereas according to Rorty, all they are are better ways of coping. Since Rorty never reconceptualizes "reality" or "correspondence" apart from the representationalist framework—choosing instead to refute it and habitually mock it—it makes no sense to him to keep talking about it. Hence, each side of the debate becomes a matter of preference. One party prefers to call something reality, whereas the other does not want to theorize about it any longer.

Another text authored by Rorty in response to Taylor can shed more light into their debate. In his chapter called "Charles Taylor on Truth" in his volume, *Truth and Progress*, Rorty says that he rejects all

forms of the correspondence theory of truth “*except* those that are so shallow and trivial as to be noncontroversial” [emphasis mine] (TP 85). This quotation provides evidence in favor of the idea that Rorty accepts ordinary uses of the concept “truth,” but not high-minded, philosophical correspondence. An example of ordinary use might be everyday conversation. It would be silly to create controversy over a person responding, “Oh, true,” after being reminded of a schedule conflict (or whatever). But also, we do not have to guess at what Rorty himself means by “noncontroversial” forms of the correspondence theory of truth because he tells us. Rorty wishes to characterize all of our true statements description-dependent. In other words, there *are no* statements that are true “in virtue of the way things are” if “in virtue of the way things are” means “*simply* in virtue of the way things are, quite apart from how we describe them” [emphasis his] (TP 86). However, everything *is* true “in virtue of the way things are” if “in virtue of the way things are” means “in virtue of the way our current descriptions of things are used” (TP 86). Hence, things can be true in the context of our descriptions, but nothing is true apart from our descriptions.

Rorty’s chief criticism of Taylor in this chapter is that Taylor wants to abandon the old correspondence theory of truth, but simultaneously argues (in texts like *Retrieving Realism*) that we are in direct contact with reality, that our scientific experiments *correspond* to “the structure of the universe.” Rorty rejects recasting the theory of correspondence in this manner:

“On my view, if we drop that distinction and that problematic [whether beliefs are made true by the world or by us], we cannot explain what we mean by ‘correspondence’ unless we posit something like ‘the world as it is in itself’ or ‘the features that a thing has intrinsically, independent of how we describe it.’ That is why I think that when the thing-in-itself goes, correspondence goes too. So I think Taylor’s attempt to keep the latter without the former is doomed to failure.” (TP 87)

Rorty thinks it makes no sense to reject the representationalist idea of “the world as it is in itself,” and yet maintain that our beliefs “correspond” to something independently of our language and practices. Taylor might respond by telling Rorty that he has not taken Heidegger’s *Zuhanden* seriously enough, and that he

should stop making language, descriptions, and higher order beliefs the paradigm cases of our interaction with the world. But still, I do not think this response would capture the dynamics underlying their debate. Rorty says that now that we have rejected the representationalist picture, we ought to likewise reject all forms of the correspondence theory of truth except noncontroversial ones, i.e. those stating that things are true “in virtue of the way our current descriptions of things are used.” Taylor also rejects the representationalist picture, but wishes to redescribe these noncontroversial cases as correspondence with “the structure of the universe.” As far as I can gather, Rorty prefers to use terms like “correspondence,” and those similar to them, according to the representationalist picture, while Taylor prefers to redefine these terms by applying them to his Heideggerian- inspired framework. As Rorty would say, each philosopher simply chooses to use a different vocabulary. One prefers to use these terms in one way, while the other prefers to use the terms in another way.

What I hope the previous paragraph demonstrates is that (1) Rorty and Taylor’s goal is the same, and (2) Rorty and Taylor’s conclusions are essentially the same, with the exception that Taylor affords his conception of the world and human practice with just as special a status as past philosophers granted to the truth of the thing-in-itself. When we use scientific methods to gain more and more control over our environment, and thereby find better ways of coping with our environment, we *really are* coming closer to understanding reality, since we have dissolved the distinction between reality and human coping. For Rorty, we should not think of our success as coming closer to reality, but as a set of human practices helping us cope. “Reality” is nothing but a tiresome shibboleth.

There is one more piece of evidence I have to offer demonstrating how Rorty and Taylor’s conclusions are virtually identical. At the end of *Retrieving Realism*, Dreyfus-Taylor argue in favor of what they call “Plural Realism.” At this point in the book, Dreyfus-Taylor place humans squarely back in the midst of our environment. All humans, through their capacity for embedded knowing, and preconceptual know-how, are in direct contact with the world. But if everyone is in direct contact with the world, how is it

the case that all humans, depending on their culture and language, do not come to the same conclusions? Even though we all are in direct contact with the world, humans disagree with one another over a wide swath of issues, which would seem to suggest that not everyone is in contact with reality. How can Dreyfus-Taylor simultaneously assert both?

For instance, in our culture, the essential nature of gold is commonly thought of as an element with an atomic number 79. The same goes for water; the essential nature of water depends upon its atomic and molecular structure: H₂O. But, other cultures, especially ancient cultures, might have believed the essential nature of gold “was sacred and so shone with divine radiance” (RR 162). Maybe they believed the essential nature of water was that it had healing properties. Whatever they believed, ancient cultures certainly did not think in terms of atomic numbers or molecular structure. But just because ancient cultures did not think in terms of atoms does not mean they were wrong about reality. We are all right. This is what makes it *plural* realism: Each culture draws upon features of its own form of life to inform the way in which they approach their environment and conceive of objects. As Hans-Georg Gadamer called it, each culture has their own “horizon” (RR 110). Each culture picks out features of the world that are uniquely significant to them. These considerations prompt Dreyfus-Taylor to confess their agreement with Rorty on this point: “Thus what gold is depends upon a culture’s practices. This tells us that the Rortyan pragmatist has got something right: one should not be a modern scientific essentialist” (RR 152). However, unlike Rorty, Dreyfus-Taylor emphasize that even though there is no single, best way for describing the structure of the universe “as it is in itself...it doesn’t follow that one has to give up a robust understanding of correspondence” (RR 151). Just because there is no best way of describing gold does not mean our descriptions do not correspond to it. Neither does it mean that we are not in direct contact with gold.

However, this is not any different from Rorty’s position. According to Rorty, what we experience in diverse cultures are differing descriptions, redescriptions, and vocabularies— differing “horizons” to use Gadamer’s term. Dreyfus-Taylor just finished explaining how different cultures develop different

perspectives by merit of these horizons, yet we are all still in contact with reality since we can all see gold, touch it, and think about it. We are still in contact with the world because the world causally interacts with us, and we with the world. This is what it means for our thoughts and descriptions to correspond with the world, not what Rorty contends. But to the contrary, Rorty accepts all of this; *he just refuses to call it correspondence*. Rorty urges that “none of us antirepresentationalists have ever doubted that most things in the universe are causally independent of us. What we question is whether they are representationally independent of us” (TP 86). By “representationally independent” Rorty means for something “to have an intrinsic feature (a feature that it has under any and every description)” (TP 86). But if this is all Rorty is questioning, he is in agreement with Dreyfus-Taylor, since Dreyfus-Taylor repudiate the existence of such a universal description—the “view from nowhere” (RR 151) description. Thus, Dreyfus-Taylor and Rorty are far more similar than they realize.

Dreyfus-Taylor and Rorty are in far more agreement than they care to admit. Just because Rorty employs the representationalist picture to demonstrate how it fails according to its own lights does not mean he is somehow a secret representationalist. Throughout their discussions, I fear that each party either merely misunderstood one another, and therefore spoke past their opponent, or decided to argue over which words to characterize their new, antirepresentationalist perspective. In the end, I find it plausible that their debate was more of a verbal dispute than a dispute over the content of their positions. Both groups chose to use a different vocabulary and tricked themselves into thinking they were in sharp disagreement.

Since the debate between Rorty and Dreyfus-Taylor is largely a verbal dispute, who presented their conclusions in a more helpful way? Whose method was better? Both groups have their own strengths. Rorty does a great job at demonstrating how the representationalist picture fails by its own lights, but articulates his position in a way that sounds reductionistic. Justification is *just* a matter of social practice; truth is *only* the consensus of human groups. But if justification is just a matter of social practice or group consensus, then human groups can develop their own social practices, their own methods of justification, and their own

truth. Each group would then possess incommensurable beliefs to other groups. Since there is no common ground between each group, the beliefs of all groups are equally valid and can only be opposed from the perspective of yet another equally valid position.

Rorty comes to this kind of conclusion when he compares Galileo to Bellarmine. The reason why we favor Galileo's "scientific" approach to astronomy and not Bellarmine's "scriptural" approach is not because Galileo's heliocentrism more accurately conforms to mind-independent facts than Bellarmine's geocentrism, but because our western consensus has since shifted to favor Galileo (*PMN* 328-331). Commenting on this passage, Paul A. Boghossian explains that calling Bellarmine's system illogical, according to Rorty,

"is just a sophisticated form of name-calling: all we're doing is expressing our preference for Galileo's system and rejecting Bellarmine's: there can be no 'objective ...standards' by virtue of which Galileo's system is better than Bellarmine's, more accurately reflective of the objective facts about justification." (*FK* 62)

Presumably, to discover these "objective standards" would mean "getting outside" our beliefs and language and so forth. It is because of passages like these that people accuse Rorty of nonrealism. Rorty makes it sound as if we can believe anything at all and no one can challenge us according to meaningful standards of rationality. We would not be able to tell the schizophrenic that there are no voices because, after all, according to the *schizophrenic's* belief system, there are. To overcome the schizophrenic's irrational beliefs we would have to establish a democratic consensus—justification can consist in nothing other than this consensus.

Rorty does a good job in arguing that past ways of conceiving of truth, knowledge, and justification were misguided, but Dreyfus-Taylor are right to point out that we do not have to be trapped in our own spaces of reason. The idea that we may not share any point of contact with people of different social practices is false given the fact that each and every one of us is already at grips with the world, possess a number of preconceptual capacities, and are embodied creatures who are all in need of nutrition, among other things.

Because of this, it is impossible, in principle, for people to be so removed from others as to have *nothing* in common with them and to possess completely incommensurate beliefs and practices. Sure, aspects of other cultures are completely alien to us, but they still inhabit the same world, contrary to Kuhn.⁸ Hence, Rorty's dissolution of the Mirror of Nature, even if warranted, does not necessarily entail his reductionistic approach, or the idea that all beliefs are equally valid. There is good reason to believe the space of reasons and space of causes are not as distinct as philosophers once held them to be.

However, although Dreyfus-Taylor hit upon a few extra insights that Rorty missed, Dreyfus-Taylor's account also contains problematic terminology. It may not be very helpful to retain weighty, metaphysically-inspired terminology like "realism," "reality," and "truth" if these words are applied to a system that is practically indistinguishable from pragmatism. Just as Rorty wishes to eliminate the distinction between representing the world and coping with it, so do Dreyfus-Taylor. Yet, Dreyfus-Taylor advocates for plural *realism*. Human groups are all at grips with the world alike, yet each possess their own unique horizons and significances, analogous to how Rorty portrayed Galileo and Bellarmine as possessing different epistemic systems. How do we derive any classic sense of realism from this?

Dreyfus-Taylor say that we do not retain any classic sense of realism, if by "realism" we mean coming up with a conception of reality that is a "view from nowhere" or understanding objects as they are *in themselves*, apart from what we think. Their conception of realism involves simultaneously affirming that we cannot make much sense of the world apart from our active interaction with it, but that this interaction itself constitutes our contact with it. Just as Heidegger decided that we could not understand what we mean by "Being" apart from our being Daseins, and just as J.L. Austin pointed out that we cannot decide

⁸ Rorty employs Thomas Kuhn's idea that people of different time periods and cultures "lived in different worlds" (*PMN* 324), and that therefore their sciences and worldviews are incommensurate with one another. Regardless of what Kuhn actually held, this is how Rorty uses Kuhn, and it is beyond the scope of my paper to decide whether or not Rorty used him correctly or misrepresented him.

whether or not the statement “France is hexagonal”⁹ is true or false apart from our own purposes in using the phrase, Dreyfus-Taylor argue that the interplay between ourselves and our environment reveals the inner structure of the universe.

But still, why call this realism? Why say that we are getting close to the underlying structure of the universe if we cannot pierce through the surface of our own social practices? Yes it is true that different cultures attribute different meanings and values to different things in different ways, but isn't it the case that just because everyone believes something to be true does not make that thing true?

In the end, Dreyfus-Taylor may have a clearer position than Rorty, but one can still accuse their claim of *Retrieving Realism* as being an empty compliment to a set of metaphysical assumptions that they themselves no longer hold. To use a favorite word of Rorty's, both groups prefer their own *vocabulary*, and it is not clear which vocabulary we ought to choose.

⁹ I thank Professor Andrew Hsu for telling me Austin's example.

Bibliography

CPR - Kant, Immanuel. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood, Cambridge, 1998.

JT - Tertaglia, James. *Rorty and the Mirror of Nature*. Routledge, 2007.

FK - Boghossian, Paul. *Fear of Knowledge: Against Relativism and Constructivism*. Oxford, 2006.

PI - Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by Elizabeth Anscombe, Wiley-Blackwell, 2009.

PMN - Rorty, Richard. *Philosophy and the Mirror of Nature*. Princeton, 1979.

RR - Dreyfus, Hubert, and Charles Taylor. *Retrieving Realism*. Harvard, 2015.

SAT - Malebranche, Nicolas. *The Search after Truth*. Translated by Thomas M. Lennon and Paul J. Olscamp, Cambridge, 1997.

TP - Rorty, Richard. "Charles Taylor on Truth." *Truth and Progress*. Cambridge, 1998. 84-97.

Promising and Its Necessary Relational Component

by Blair Kimbal

A promise is a powerful communicative tool a speaker uses to assure a listener of future action. Unlike instances of merely stating your intentions, a promise requires that you follow through on the action you promise and does not allow you to change your intentions for action. A promise can be given by a promisor to a promisee allowing them to develop special, reliable expectations and is thought to be morally binding. In the following paper, I'll explore two theories that strive to explain what should compel us to keep our promises and why breaking those promises is wrong. John Rawls presents a conventionalist theory that appeals to the principle of fairness for the binding obligation of a promise. Thomas Scanlon presents an expectation-based theory that cultivating and then breaking expectations in a promisee is wrong. In this paper I will present both Rawls and Scanlon's accounts for what binds a promisor to keep her promise and why breaking such a promise is wrong. However, Rawls and Scanlon's accounts are both fundamentally unsatisfying because neither captures the deeply personal nature of promise making and breaking. I will ultimately propose an additional, relational component of promises that contributes to their morally binding nature.

Rawls begins his conventionalist account by identifying the social advantages of promising, which is a mutually advantageous agreement of cooperation (Rawls, pg.3). A promise is, in its nature, made between two persons. Although none of us sat down with every person we've ever made a promise to and agreed to engage in promissory acts reliably, we all have a general understand of what can be expected from a promise. Everyone who voluntarily engages in the act of accepting promises can reliably depend on what was promised, and anyone who voluntarily engages in the act of making a promise benefits from being able to be taken at her word. When a promise is made, the promisor generally need not give further proof to substantiate the seriousness of her promise. Rather than speak at length about your reliability for performing

an intended action and assuring another person that they can expect you to perform that action and that they should have a complaint if you do not perform the action, the word “promise” is like a social shorthand to ensure your promisee of these things. A promise can put a promisee at ease for future action and can save a promisor the time and energy it might take to assure a listener of commitment to action.

An intention to perform an action may require further evidence that the action will actually be performed, but there are social reasons to believe that a promise to perform an action needs no further evidence. If I state that I intend to walk your dog, you may believe that I will perform the action stated. But, you may be justified after the statement of my intention, to still wonder if I actually will walk your dog. Simply having the intention to walk your dog leaves me the opportunity to change my mind to not walk your dog, and it seems you could not make a moral claim against me. On the other hand, if I *promise* to walk your dog, you have no reason to wonder if I actually will walk your dog, barring some unforeseen circumstance, because of the nature of the word promise. In our social culture, a promise binds me to perform the promised action. If, after promising to walk your dog I change my mind, you could have some sort of moral claim against me that you would not have had if I only stated an intention.

If after not following through on a simple intention to walk your dog, you bring a claim against me, it would be something like “how dare you change your mind!”. Most of us maintain a right to change our mind in activities we voluntarily elect to do, so it seems your claim against me is not very strong. Relevant explanations for changing my mind could be as simple as “I intended to walk your dog but I went to the aquarium instead”. This seems to be a sufficient explanation because I have the ability to change my intentions. Now if instead I had promised to walk your dog, your claim against me would be “how dare you break a promise!”, and my explanation of going to the aquarium instead no longer seems sufficient. Your claim against me is a much stronger one in this case because it is a moral claim. A broken promise entails some sort of moral wrong that a changed intention does not. At the same time, promise-making allows assurance in a way that mere intention stating does not.

The mutual benefit that comes from being able to make a promise without incessant substantiation and accept a promise without worrying about completion of the promised action, lends itself to the creation of a social institution of promising. A social institution is a system of behavioral patterns that are deeply woven into a society. The simple fact that I can put someone at ease when I make a promise to do something shows this social institution in practice. Society in its entirety never sat down to agree to what promising entails, but it is apparent that promising patterns occur reliably in our society. Whenever someone takes advantage of the benefits of a social institution, such as promising, she must follow the rules and uphold the principles of that institution (Rawls, pg. 1). This is Rawls's principle of fairness. This principle obligates those who benefit from promissory relationships to follow through on their promises, and thus uphold the rules of the institution of promising. The principle of fairness is a necessary principle for maintaining any social institution. Because the institution is social, it is both created and maintained by the people participating in it. In order to preserve any social institution, including that of promising, the people participating in it must consistently follow its rules. The principle of fairness ensures that in participating in the institution of promising, someone is obligated to uphold the rules that allow the institution to continue to function. This principle develops our obligation not to break our promises on a conventionalist view. If someone did not uphold their promises, but still sought to benefit from the social institution of promising, they would be doing a sort of *free-riding*: benefiting from an institution, but not doing their fair share to maintain that very institution. Free-riding violates the principle of fairness and does not uphold the social institution of promising.

Although the principle of fairness obligates every individual promisor, the problem with Rawls's conventionalist theory is that it does not capture the personal, relational aspect of promising. On Rawls's account, you are bound to keep your promises by the principle of fairness that applies universally to all participants of the social institution of promising. But when you break a promise to someone, your promisee has some sort of special complaint against you that everyone participating in the institution of promising

does not have. If the wrong of breaking a promise is that I violate the principle of fairness and take a free ride on the social institution of promising, then every person participating in that institution would have the exact same claim against me. This implies that by breaking a promise to *you*, my cashier at the grocery store who also participates in the institution of promising would hold the same claim against me that you have. Why should anyone participating in the institution of promising get to hold the same morally relevant claim against me as the person I directly engaged in the promise? A theory of the wrong of promissory breach that exposes the promise-breaker to one single universal complaint does not sound like it could be right.

The conventionalist view does not match up with the actual complaints we make against promise-breakers. If I were to break my promise to you, you would likely not be outraged that I violated the principle of fairness; you would be outraged because I violated *you* in some sort of personal way. If I promised to turn in your end-of-term paper to your professor's mailbox before its deadline, and then did not turn the paper in, you would not be frustrated that I was free riding and benefitting from the social institution of promising while not upholding its rules. You would be angry with me for violating you in some personal, relational way that comes with a broken promise. I would have violated trust that you put in me.

This is a problem for Rawls's view because if the moral obligation and moral wrong of promissory breach reside in a social convention of promising and in the principle of fairness, why would a promisee feel personally wronged? The promisee should want to shame the promisor for breaking the principle of fairness on Rawls's account, but instead they are hurt in a far more personal way. Rawls's view cannot explain this.

Scanlon replies directly to Rawls's conventionalist view by imagining cases where no social institutions exist, yet it still is wrong to break a promise. If you are still bound to keep your promises without a social convention in place, the convention cannot be the reason you are bound to keep your promise (Scanlon, pg. 215). To show this, Scanlon brings up reasonable moral principles that do not rely on a preexisting social convention.

Scanlon imagines several moral principles that he believes are generally accepted and practiced. These principles are general and reasonable in the sense that no one could reasonably reject them. One of these principles addresses the wrong of intentionally manipulating someone into doing something because they believe you will do something in return, when you have no intention to actually do what they expect (Scanlon, pg. 203). Scanlon says that no one would reject such a principle. It seems like a generally morally good thing to accept a principle that disallows manipulation, and it being morally good does not rely on a social institution.

Scanlon is a contractualist in that he believes reasonable moral principles are in fact the principles that morally bind us. For something to be a reasonable moral principle, it must be one which no rational agent could reasonably object. In the absence of a reasonable rejection, Scanlon thinks that all moral agents could “consent” to these being binding principles that regulate our actions. This is the sense in which agreeable principles constitute something like a moral contract, since Scanlon believes it's one that all agents would agree to. As rational beings, we make *the right choices* when it comes to decisions in a social sphere. This includes what sorts of moral principles seem reasonable enough not to reject. The moral principles we are able to agree on are the ones which obligate us.

Other principles he addresses involve practicing due care not to lead someone to develop false expectations or taking a certain course of action to prevent or compensate for any loss that would result from these expectations. He calls on general reasonability to make a case for these principles being ones to accept. If you were using a neighbor's kitchen in order to bake some bread, and led them to the expectation that you have turned off their oven, when you actually have not, you have led your neighbor to have false expectations (Scanlon, pg. 204). It is your duty, according to Scanlon, to warn them that their oven is still on before anything bad happens in their kitchen, or, if something bad does happen, to pay for the damage. These obligations exist to Scanlon because most people can agree to them.

Scanlon says that if you can accept these general moral principles as plausible obligations, you would also reasonably want a principle that requires performance of action rather than just compensation for loss. This is the principle that is used to explain why we are obligated to keep our promises. This principle says that you are obligated to perform an action that you have led someone to expect you to perform, especially if you have assured them that the action will be performed. It is a general moral principle like the others but it instead requires performance, not just compensation (Scanlon, pg. 208). Because most people will agree to these sorts of general moral principles, we have come to expect the sort of behavior allowed by these principles, and feel violated by behavior that is disallowed by them. When it comes to making a promise, Scanlon holds that because of these general moral principles, people can reliably expect promises to be executed.

On this view, the wrong of breaking a promise comes from what we owe someone when we lead her to form expectations about our future actions (Scanlon, pg. 200). In creating expectations with a promise, the promisor now morally owes their promisee a following through on the promise. Cultivating expectations for future action in a promisee through a promise gives the promisor a duty to uphold those expectations. Scanlon hopes to demonstrate that if you can plausibly agree to moral principles that obligate us to compensate when we cannot follow through on the expectations we have developed, you would agree to a moral principle that makes it wrong not to follow through on our promises.

What is important to Scanlon's account is that there is value in the assurance that your expectations will be fulfilled. Under the domain of this moral principle, the promisee has reason to rely on performance of the promised action rather than just to expect equal compensation for loss. In a case where your mother *promises* to pick you up from the airport after you've spent a long time away from home, it is reasonable for you to expect to be picked up by your mother. If she were to instead send a driver to get you, or reimburse you for the price of a taxi home, you might think what she has provided as an alternative is not quite the same as if she had fulfilled her promise. In this case, finding a similar solution or compensation does not

meet your expectations for action. You may have really valued that your mother pick you up, and to not follow through on the action, despite providing something to produce the same outcome, seems to fall short of what a promise entails. In making a promise on Scanlon's account, I commit to an action for you and do not commit to anything less.

This account of the wrong of breaking a promise succeeds in the relational aspect better than Rawls's account does. It seems to tie the promisor to the promisee directly, rather than just tie the promisor to a social institution. In making a promise, I owe it to you to follow through and uphold that promise. I am obligated to fulfill my promise by the expectations in *you* and am therefore bound to *you*. Scanlon seems to begin to capture something important about the personal nature of promising.

The problem with Scanlon's expectation theory is it seems to claim that insofar as you are able to develop an expectation in someone, you are able to make her a promise. However, it seems you can develop expectations without actually making a promise. Consider the following case: You approach a stranger on the sidewalk and attempt to convince them that if they stay where they are, you will bring them money. This stranger is thrown off by your unsolicited promise and no matter how much you insist that your promise is one to rely on, one might not be surprised that they have trouble accepting your promise or believing that your declared intentions will be fulfilled. Without any reason besides your dogged insistence, the stranger has no rationale for expecting your promise. No matter how hard you try, you may not be able to engage this stranger enough for them to expect your promise to occur.

If somehow, after hard work and plenty of convincing, you were able to instill the expectation in them, it would seem you have done more persuasion than the offering of an invitation to rely on your promise. They might expect you to bring them money, but only because you went to such an effort to convince them it would happen. Although their expectation exists, hardly anyone would claim that you are now *morally* obligated to bring them money! It certainly would be odd to insist on something so persistently and then not follow through, but there really isn't much morally wrong with this sort of behavior. This is a

problem for Scanlon because the only reason they have to expect your promise is your convincing. If your obligation to fulfill the promise comes from their expectation, and their expectation comes from your convincing, it seems almost like you have forced your own obligation into existence through persuasion.

Moral obligations seem like the sorts of things we cannot conjure up on our own. If expectations are the basis for moral obligations, it seems like we could create the moral obligation to do almost anything! If I am very gullible and you are able to convince me that you are joining the circus, to the point where I truly do expect it, it seems wrong to think that you are now suddenly obligated to join the circus. You may have wrongly deceived me, but there is no requirement that you follow through on my expectation. Expectations on their own may not be a proper basis for moral obligation, because they cannot explain the strong bind we feel with someone in a promissory relationship that seems to make a promisor bound morally to their promisee. Something must exist outside the expectation that would allow us to rely on the moral force of promising.

It seems to me there is a deeply personal wrong of promise breaking that does not follow from violating a social convention, or simply not fulfilling another person's expectations. A promise may not only be an invitation for a promisee to develop special reliable expectations, but there may also be a component of trust necessary for a promissory relationship. The wrong of promise-breaking may come from the breaking of ties of trust that linked two individuals before the promise was made.

Trust is a sort of relation that must be developed rather than merely established. Trust is not created, but cultivated with time and consistency. In order to trust someone, you might need evidence that they have upheld their promises in the past or that they have been reliable in other contexts without promises. In developing a relationship with someone you begin to rely on them and develop patterns and consistency. Through demonstrating your own consistency in speech and in action, you are able to align your goals and methods of communication with that other person. These patterns are what enable you to begin to know someone. When you begin to know someone, you're more able to rely on that person, and this reliance

compounds on itself. The more you know someone, the more you are able to rely on them, and the stronger the ties of trust you have together are. You would have no reason to *trust* a promisor you just met, even if you did *expect* them to follow through. Expectations can be created but trust needs to be cultivated.

When a stranger creates expectations in you and does not follow through on them, your expectations have certainly not been met, but the betrayal you feel is far less than if a best friend of yours did the same. This feeling of betrayal might not be the best measure of moral failure, but it seems that your friend has failed you in a different way than the stranger has failed you. Your friend has betrayed the trust you two cultivated together and damaged your relationship.

There may be a need for some outside connections, or ties of trust, to exist between two people before a promise can be made. You can insist on expectations to be formed in a stranger, but for a promise to be made, the promisee must trust you. In breaking a stranger's expectation, you let them down, but in breaking a promise to a friend you damage your ties of trust that built the relationship. Ties of trust *must* exist between a promisee and a promisor for a promise to be made rather than simple expectations for action.

It seems compelling to uphold a promise for another individual and the relationship you have built with that individual, rather than just for the social institution of promising or for expectations for action. My account offers the additional reason to keep your promises by maintaining ties of trust and a relationship that has been built. Promising is innately relational and you do not just owe it to someone's expectations to uphold your promises, you owe it to them and the ties of trust between you.

A major problem with my account is that it seems like, in some circumstances, strangers might be able to make promises to each other. Imagine two people who are naturally very willing to engage strangers. When these two people meet as strangers to one another, they may decide to enter into a promissory relationship that the first man will drive the second man to his car in exchange for the beach umbrella that is in that car. It seems like if they arrived at the car and the second man refused to give the first man his beach umbrella, he would be breaking a promise and doing something morally wrong. There are no ties of trust in

this relationship, yet the second man certainly has done something wrong. The first man trusted the second man without any previous ties of trust.

By not upholding a promise, a complete stranger may be able to wrong a promisee. This is a problem for my view because the large moral failure doesn't seem to rely on a previous relationship. There are no previous ties of trust between the two strangers to break; yet the promisee can still be wronged. I'd like to argue that there never was a promissory relationship between these two men; there was only a general obligation for action that can be generated by giving your word.

We may have a baseline of obligation when we act to cultivate expectations in others. Similar to what Scanlon imagines, it seems generally wrong to go back on your word or insist upon something and then easily allow it to fall by the wayside, no matter who you are engaging with. If Sue is moving into a new apartment building and one of her neighbors, on sees her unloading a moving van, says he will come down to help her unload the truck, Sue is reasonable in expecting him to return. If after she has finished unloading her belongings the new neighbor has not returned, Sue may be confused and had her expectations let down, but seems not to have had a promise broken. Her neighbor may have violated a general moral rule of not going back on your word when you give it, but he could not have committed a more serious moral wrong similar to promise breaking. It would sound too strong to say that her neighbor, a stranger to her until that moment, had promised to help her move in and then subsequently broken that promise. This suggests that we do have a baseline of obligation to act when we cultivate expectations in others, but this obligation is more similar to that of a statement of intention. In intending to do something, we are morally required to provide some sort of explanation for why we cannot follow through when we cannot follow through, but this not following through on intentions is a far less significant moral wrong than breaking a promise.

In having a baseline obligation to one another to uphold our word when we give it, we allow for a kind of social stability that would be impossible without it, though this baseline obligation does not appear to me to be one that allows the development of a promise. In order for a promise to be made, a bond to tie the

expectations of one person to the action of the other, something must exist for those expectations to be founded on. Justified expectations for promissory action rely on the existence of trust between two people. From a baseline obligation we have to one another, individuals are able to develop relationships. From these relationships stems trust and with trust may come a stronger relationship that allows individuals to make promises.

Although you may be able to create expectations in strangers and break expectations in strangers with no previous ties of trust, ties of trust can enhance the baseline obligation you have to someone. Trust allows someone to build upon this baseline obligation to follow through by providing reasons to follow through besides the baseline obligation. Once trust exists in a relationship, maintaining the trust and continuing the relationship provide extra moral reasons, on top of the baseline obligation, to uphold a promise. When you make a promise to a friend, you are not only taking on the obligation to fulfill their expectations, you also take on the obligation to maintain the ties of trust between you. It seems quite possible that when a promise to a friend is broken, more than just their expectations are violated; something else is deeply disrupted between the promisee and the promisor. The relationship required to make a promise ensures that trust exists between both people. When a promise is broken, expectations are let down and trust may be broken as well. It may be worse to break a promise to a close friend than to a stranger because of the disruption of personal ties of trust. Ties of trust bind you morally to others and make keeping promises a deeply personal moral obligation.

Bibliography

Rawls, John. "The Arguments for the Principle of Fairness." *A Theory of Justice*.

Cambridge, MA: Belknap of Harvard UP, 1971. N. pag. Print.

Scanlon, Thomas. "Promises and Practices." (n.d.): n. pag. Web.

The ‘Now What?’ Problem for Moral Error Theory: Substitutionism vs. Revisionary Moral Expressivism

by Shayan Koeksal

Moral error theory’s thesis states that (a) moral judgments properly understood are truth-apt, and yet (b) there is no moral truth. From this it follows that all moral judgments properly understood are necessarily erroneous. If this is correct, or if one is convinced by the arguments for moral error theory, then this puts the moral error theorist in an uncertain situation. If she does not desire to commit crime or cause chaos, what should she *now* do in order to succeed and flourish “morally” in society? How should her second-order metaethical beliefs inform her first-order thoughts and actions in a coherent and rational way? What should she think, say, and do? This is the ‘Now What?’ Problem for moral error theory.¹⁰ In this paper, I canvass the dialectic of the debate spurred by the problem and briefly survey previous attempts at solving the problem (abolitionism, conservatism, and fictionalism)¹¹ before focusing in on two unchallenged recent solutions (substitutionism and revisionary moral expressivism) that claim to avoid the issues of the former attempts. Finally, after arguing against revisionary moral expressivism, I spend the rest of the paper developing my novel substitutionist solution.

I. Introduction

In this paper, I present the ‘Now What?’ Problem in detail (§II), make preliminary clarifications (§III), briefly survey previously proposed solutions (§IV), and critically evaluate two unchallenged recent solutions: substitutionism (§V) and revisionary moral expressivism (§VI). Next, after arguing that revisionary moral expressivism succumbs to the Frege-Geach problem, I motivate the intuitive plausibility

¹⁰ The term “the ‘Now What?’ Problem” was coined by Lutz (2014).

¹¹ I should note that I am only surveying these three solutions to give the reader a better idea of how the dialectic has gone previous to Lutz’s (2014) substitutionism and Svoboda’s (2015) revisionary moral expressivism. Both Lutz and Svoboda take abolitionism, conservatism, and fictionalism to incur too many costs, thus, why I spend the majority of the paper tackling their respective positive contributions.

of substitutionism (§VII). I then spend the rest of the paper developing my novel substitutionist solution (§VIII) before, finally, concluding the paper (§IX).¹²

II. The ‘Now What?’ Problem in Detail

Before introducing the problem, I must explain what moral error theory is. Moral error theory, in the form of a general argument (which is usually credited to J.L. Mackie),¹³ has a conceptual claim and an ontological claim. The conceptual claim is that moral discourse has non-negotiable commitments to properties that are objectively prescriptive and categorical.¹⁴ The ontological claim is that there are no such properties. Put informally:

1. If there is moral truth, then it has non-negotiable commitments to objectively prescriptive and categorical properties (conceptual claim).
2. Nothing is objectively prescriptive and categorical (ontological claim).
3. There is no moral truth (from 1 and 2)

The problem that arises for the moral error theorist is that there is no obvious solution as to how her second-order metaethical beliefs (moral error theory) inform her first-order thoughts and actions. There is no clear implication of what she should do from what she believes. An adequate answer to this issue is important because, on one hand, it offers the distressed error theorist a solution to her first-order discourse and, on the other hand, shows that error theory does not entail or countenance heinous criminality, genocidal maniacs, or the rise of thousands of Jeffrey Dahmers. It has been brought up before that, to non-error theorists, it seems like accepting error theory’s thesis precludes any kind of important “moral” normativity in

¹² Clarificatory work (pp. 2-5), exegesis (pp. 6-12), novel critical evaluation and development (pp. 13-22).

¹³ Mackie (1977). However, the conceptual and ontological claim framework I am using has been taken from Joyce (2001, p. 13).

¹⁴ This is not an uncontroversial premise. Whether or not categorical prescription is a non-negotiable commitment has sparked interesting discussion. See the back-and-forth between Finlay (2008) and (2011) and Joyce (2011) and (2012).

one's life.¹⁵ A good solution to the problem will show that these initial concerns are mistaken and will ultimately make error theory a more palatable thesis.

The 'Now What?' Problem originates from Mackie, when he first introduced moral error theory, arguing "one could be a second order moral sceptic without being a first order one" (Mackie 1977, p. 16). Since then, there have been various attempts at resolving the disharmony of being an error theorist but still wanting to succeed and flourish "morally" in society.¹⁶ In section four, I survey abolitionism, conservationism, and fictionalism as a way to show how these solutions fill the conceptual space and have led to the two I later evaluate: substitutionism and revisionary moral expressivism. I do not discuss the former three solutions in any great detail, as these have been discussed at length already.¹⁷ My main focus this paper is on the latter two as there is nothing in the current literature that I could find that challenges these solutions and, also, because their proponents claim them to be the best solution to the 'Now What?' Problem.

III. Preliminary Clarifications

With the problem now better specified, there are still some concerns that the reader might have. This section clarifies any immediate worries and makes important distinctions.

The first worry that might strike the reader regards how attempting to find a solution to the problem is not a self-defeating task for the error theorist. They might ask, "If you hold that there is no categorical prescription, is it not undermining that this project is attempting to find something like it?" This brings up an important distinction between categorical and hypothetical imperatives. The issue moral error theorists have with morality is their belief that it commits us to the existence of objective, categorical normativity.¹⁸

Categorical imperatives are supposed to provide reason-giving normative force regardless of one's desires.

¹⁵ Lutz (2014, p. 370) speaks to how error theory is often taken as being the death of any kind of "fulfilling normative life."

¹⁶ Of course, stipulating that the error theorist does not want to go on and use their second-order beliefs to justify heinous crimes.

¹⁷ If, however, you are interested in these solutions, then see Olson (2011) and Olson (2015, p. 190-198) for conservationism, Hinckfuss (1987) and Garner (2007) for abolitionism, and Joyce (2001, p. 175-231) Nolan, Restall, West (2005) for fictionalism.

¹⁸ This is the property that Mackie labeled as "utterly different than anything else in the universe" (1977, p. 38).

Hypothetical imperatives, on the other hand, are contingent on one's desires¹⁹ and, generally, moral error theorists do not have issue with this kind of normativity.²⁰

To further clarify this distinction, imagine a world where the only reason anyone ever goes to McDonald's is because they desire the type of hamburger served there. It makes sense that, if I have the desire to eat that hamburger, then I have reason to go to McDonald's. The desire provides a kind of normative reason-giving force. This is the kind of hypothetical normativity that I am assuming is unproblematic in this paper. However, the kind that error theorists normally take issue with is the kind that gives reason independent of desire. If I do not desire a McDonald's hamburger, for whatever reason, then I have no reason to go to McDonald's. Analogously, in morality, if I know that murder is wrong, but I do not desire to act morally, then my knowledge of that proposition provides me no hypothetical reason not to murder. The hypothetical imperative at work in the background of this paper is for those who desire to succeed and flourish "morally" in society even with their error theoretic inclinations. The self-undermining objection is avoided by searching for strong hypothetical desires, not categorical ones.²¹

Also, the type of error theory that will be referenced when error theory is brought up will be moral error theory, which concerns those that have issues with categorical normativity in morality. There are global error theorists who have written on this topic,²² but I will only be concerned with local error theory that concerns the moral realm in this paper.

The last clarification regards the term "Now" in the 'Now What?' Problem. This is not the 'In the distant future problem, what?' problem. A good solution to the 'Now What?' Problem should take into account concerns that might get overlooked, or even be considered unimportant, if one looks too far into the future. How a solution allows the error theorist to communicate with other philosophers and non-

¹⁹ Kant (1785, pg. 74)

²⁰ Streumer (2013) and Köhler and Ridge (2013), however, have global normative concerns.

²¹ The error theorist is not alone in believing hypothetical imperatives can still command strong reason-giving force. Foot (1972) argues that, even if there are no categorical imperatives, people still care about hypotheticals such as justice and liberty.

²² Köhler and Ridge (2013)

philosophers (without stipulating a kingdom of error theorists) is an important issue. Avoiding awkward discourse and not needing society-reforming stipulations are two points central to a good solution to the ‘Now What?’ Problem.

IV. The Dialectic till Now: Abolitionism, Conservation, and Fictionalism

In this section, I briefly survey past attempts at resolving the problem in order to give the reader a better idea of the dialectic leading to substitutionism and revisionary moral expressivism. I explain what, canonically, are taken to be each solution’s advantages and disadvantages. This should provide an idea of what is, and what is not, valuable in a given solution.

Abolitionism, usually credited to Hinckfuss (1987) and Garner (2007), states that the solution to the ‘Now What?’ Problem lies in giving up moral language and action based on the presumption of moral truth. The abolitionist discards moral discourse in the same way that, upon realizing its systematic falsity, one discards all astrological discourse.²³ Further, Garner argues that this elimination of moral discourse will be “no more difficult than cutting back on swearing” (2007, p. 512).

The obvious benefit of abolitionism is that it avoids moral error by eliminating moral-based action and talk.²⁴ However, there are many concerns with abolitionism that make it to be an implausible solution. First, the concerns over moral talk and action being used as an oppressive tool are overblown.²⁵ Second, abolitionism might actually be psychologically impossible and attempting to rid us of a mode of discourse that is likely evolutionarily ingrained is very unlike giving up swearing.²⁶ Third, the abolitionist is “unable to engage with others in moral discourse” because any attempt at entering moral debate will be, at best,

²³ The abolitionist’s justification for eliminating moral discourse is not just based in morality being error-laden. Both Hinckfuss and Garner hold that morality creates more problems than it solves and actually “inflames disputes and makes compromise difficult, preserves unfair arrangements and facilitates the misuses of power, and makes global war possible” (2007, p. 502).

²⁴ Also, the type of practical rationality Garner wants to replace morality with seems promising in its avoidance of the issues that he believes plague moral arguments (2007, p. 507).

²⁵ Nolan, Restall, and West respond to abolitionism by arguing that, in fact, “Moral discourse is extraordinarily useful. Morality plays an important social role in coordinating attitude and in regulating interpersonal relations” (2005, p. 507)

²⁶ Lutz (2014, p. 357)

awkward.²⁷²⁸ This might not be a problem given a kingdom of abolitionists using this kind of rational calculation, but, given that this is the ‘*Now What?*’ Problem, it is a cost that abolitionists incur due to the first-order moral language an error theorist is likely to encounter.

Conservationism, credited to Olson (2014), recommends biting the error-laden bullet and endorsing moral talk and action. The conservationist’s prescription is that one should compartmentalize, based on context, when they literally believe the error theory.²⁹ Olson argues that this is possible by utilizing occurrent and dispositional beliefs.³⁰ For example, when the conservationist is discussing the wrongness of a heinous crime, then she believes that occurrently while believing the error theory dispositionally. Lastly, Olson contends that this kind of engaging with moral talk will serve as motivation to avoid acting “immorally.” If a conservationist says murder is wrong, and engages with this kind of moralized thought, then she will be less likely to murder.

Though Olson takes his conservationism to be able to hold on to the benefits of moral discourse, there have been various criticisms that put conservationism in a bad light. The first, and most obvious, criticism is that conservationism endorses systematic moral error (Svoboda 2015, p. 15).³¹ Second, is that occurrent and dispositional beliefs are still full beliefs.³² This results in a kind of irrational Orwellian doublethink where a conservationist both believes and disbelieves a proposition such as “murder is wrong.” Finally, if we stipulate that occurrent and dispositional beliefs are not full beliefs, then conservationist

²⁷ Lutz (2014, p. 358)

²⁸ We can imagine an abolitionist entering a debate on whether or not killing is morally right and responding that “my rational calculations make it so killing is often not to my personal benefit.” This kind of response will only work to derail the debate and make any kind of resolution very difficult.

²⁹ For an example, Olson recommends maintaining “moral belief in morally engaged and everyday contexts and [reserving] attendance to the belief that moral error theory is true to detached and critical contexts, such as the philosophy seminar room” (2014, pp.191-192).

³⁰ He uses the example of a lying politician to show how one can know someone is lying, but be so seduced by their speech that, in the moment, they take the speech to be true (Olson 2014, pp. 192-193). Also, the occurrent and dispositional belief distinction is analogous to beliefs being at the front versus the back of one’s mind. For example, I have the occurrent belief that I am typing right now and the dispositional belief that I prefer strawberries to bananas.

³¹ Further, Garner believes this “undermines our epistemology by promoting a culture of deception” (2007, p. 499). He actually states this in response to fictionalism, but the point is as relevant here.

³² Lutz (2014, p.355) and Svoboda (2015, p.12)

reason-giving force is lost. A conservationist saying, “murder is wrong” (if not fully believed) will do nothing to motivate her actions, much as a person who has taken horoscopes to be nonsense saying her own horoscope out loud. This person has no reason to act on any prescriptions mandated by the horoscope.

Fictionalism, credited to Joyce (2001), recommends that the best solution to the ‘Now What?’ Problem lies in taking a fictional attitude towards moral propositions.³³ The fictionalist holds on to surface cognitivist assertion, but uses an implicit operator so that “murder is wrong” really means “murder is wrong according to [something like] the best story of morality” (Joyce 2001, p. 195). This avoids falsity since the implicit operator allows the fictionalist to take a fictive attitude towards seemingly genuine assertions.³⁴ Furthermore, the fictionalist contends that taking a fictional attitude will provide motivation to act in a “moral” way the same way watching the horror movie *The Blair Witch Project* motivates one to avoid camping in the woods (Jaquet and Naar 2016, p. 199). Lastly, fictionalism allows the fictionalist to engage in first-order discourse.

Fictionalism does avoid the issue of moral error, but is unable to provide a good prescription that motivates moral action and reasoning. First, it is irrational to act on a fiction.³⁵ If one watches the most recent *Harry Potter* film and sees that the title character is in imminent danger, it would be delusional to attempt to fly to the campus of Hogwarts to warn him. Second, it is implausible that a fictional attitude could provide real motivation to act “morally.” It does not seem likely that the fictionalist will be motivated to act on something that she takes to be make-believe. The same reason that no one goes unicorn hunting is the same reason why the fictionalist account does not provide strong reason-giving force. This make-believe approach to morality leads Lutz to contend that it “[trivializes] our deepest commitments” (2014, p. 361).

³³ Fictionalism can be further separated into force and content versions, which Naar and Jaquet (2016) discuss, but for the sake of this paper’s length I do not delve into this.

³⁴ For example, saying, “Ron Weasley has orange hair” is literally false because no such person exists, but if one takes a fictional attitude (and assumes an implicit operator), then the statement “According to the Harry Potter series, Ron Weasley’s hair is orange” is true.

³⁵ Lutz (2104, p. 360)

In sum, abolitionism avoids moral error, but is founded on exaggerated premises, might be psychologically impossible, and leaves the abolitionist unable to engage in first-order moral discourse. Conservationism holds onto some of the benefits of moral discourse (such as the ability to engage in it), but endorses moral error and Orwellian doublethink, and provides weak motivation. Lastly, fictionalism avoids moral error and holds some of the benefits of first-order moral discourse, but advocates irrationality in moral action and reasoning.³⁶ The advantages and disadvantages of each of these solutions should be taken into account when I present substitutionism and revisionary moral expressivism in the next sections.

V. Substitutionism

Substitutionism, credited to Lutz (2014), argues that we should “adopt a fully-committed attitude to some propositions that are not moral propositions, and use those full commitments to fill the hole in our normative life” (2014, p. 362). I will expand on this in a moment, but it is important to explain Lutz’s reasoning. After surveying abolitionism, conservationism, and fictionalism, Lutz contends that an ideal solution to the ‘Now What?’ Problem should meet three criteria. The first is that the solution should not endorse any positive moral beliefs. The second criterion is that the ideal solution should allow the error theorist to act morally³⁷ and use moral language when appropriate. The third criterion is that the action and language “cannot be based on the acceptance of a moral proposition.”³⁸

Next, Lutz explains how non-moral propositions are supposed to substitute for moral ones by introducing non-negotiable and negotiable commitments, defective commitments, and a salvaged concept. Briefly, non-negotiable commitments are those commitments that are necessary for something to be what it is, whereas negotiable commitments are ones that can be true contingently. For example, a triangle has the non-negotiable commitment of having three sides. Something cannot be a triangle without have three sides.

³⁶ Again, this exegesis is not exhaustive but presents a sufficient picture of the issues with the aforementioned solutions.

³⁷ I address this when I evaluate this solution, but this is more “morally” than morally. It is what is generally considered moral, but is actually non-moral substitutionist action.

³⁸ These three criteria can be found in Lutz (2014, p. 362).

A negotiable commitment might be the color or size of the triangle. Applying this to morality, Lutz argues that error theorists should replace (or substitute) first-order moral discourse with a non-moral concept that captures everything valued of morality. Of course, this is not totally possible given that error theorists take categorical prescription to be a non-negotiable commitment of morality. Lutz calls this categorical commitment the *defective* commitment. By ridding of the defective commitment, the error theorist is no longer making reference to moral truths (since, for the error theorist, moral truth requires categorical prescription).

The next step is to build up a non-moral framework that captures the benefits of moral discourse without the issues. Lutz does this via Richard Brandt's *salvaged* concept.³⁹ The salvaged concept functions as morality-minus-categorical-prescription (which is still distinctly non-moral). It will encompass all the other non-negotiable and negotiable commitments that are valued of morality. The non-negotiable commitment Lutz has in mind is allowing for "approval or disapproval towards someone who does or does not do what they have most reason to do," but this list can likely be extended to include things such as making allowances for moral reasoning and disagreement, providing strong motivation to act accordingly, etc. (2014, p. 364). The negotiable commitment could include what provides a "moral" reason. This cluster of non-negotiable and negotiable commitment makes up the salvaged non-moral concept that substitutes in for moral discourse. Simply put, the salvaged concept works as a "not morality (and does not commit us to spooky objective prescription), but close enough" operator.

Substitutionism also allows flexibility in what one substitutes in. Lutz defends a Humean breed of substitutionism, which I examine later, but now I will exemplify how a certain naturalist reductionist substitutionist (who reduces goodness to pleasure) would work. When in first-order discussion, she would say, "murder is wrong," but would not believe the proposition as such and would instead substitute in (that what she really means is) "in the case of murder, pleasure is not being maximized." This has the kind of

³⁹ Brandt (1998, pp. 10-16)

implicit operator that fictionalism offers, but without the irrationality, and lack of motivation, since this person very much believes that naturalist reductionism offers the best substitutionist salvaged concept of morality and does not take it to be fictional. It seems that the substitutionist's salvaged concept offers a very attractive way to fill in the holes that error theorists face.

VI. Revisionary Moral Expressivism

Revisionary Moral Expressivism, credited to Svoboda (2015), argues that the error theorist should adopt expressivism in light of the 'Now What?' Problem. More specifically, he means that we should adopt "expressivist, non-cognitivist moral discourse and [jettison] any cognitivist moral discourse and judgment" (Svoboda 2015, p. 15). As a result, moral judgments will be changed into desire-like attitudes and the subsequent moral talk will be expressions of these attitudes. Svoboda proposes, for example, changing "murder is wrong" to expressions like "boo murder" or "don't murder," but also states that such a drastic change might be unnecessary and that cognitivist assertion could just refer to one's expression of their disapproval. Further, Svoboda takes non-cognitivist utterances to be in the business of delivering moral judgments. In sum, the revisionary moral expressivist should get rid of their moral beliefs and replace them with non-cognitive attitudes.⁴⁰

Similar to Lutz's criteria, Svoboda evaluates his solution according to how well it avoids moral error, and holds onto the intrapersonal and interpersonal benefits of morality. The way revisionary moral expressivism avoids moral error is by either avoiding use of truth-apt beliefs or having truth-apt moral language actually express non-cognitivist desire-like attitudes. Utterances such as, "boo murder" or "don't murder" do not have a truth-value and, thus, are incapable of being false. Intrapersonally, Svoboda holds that revisionary moral expressivists will have motivation to act in accordance with their desire-like attitudes. If one expresses the moral judgment that they disapprove of lying, then it is straightforward that this will motivate the expressivist not to lie much as disapproving of the taste of salmon will motivate one not to eat

⁴⁰ Svoboda (2015, p. 17)

the fish (Svoboda 2015, p. 20). Interpersonally, Svoboda argues (citing Stevenson)⁴¹ that because expressivists are not putting forward beliefs when arguing, they can see disagreements dealing with differences in attitude. Further, the revisionary moral expressivist can account for moral reasoning by using pragmatic consistency.⁴²

Again, on the surface, it looks like the revisionary moral expressivist has a good case. In the next section, I evaluate both Lutz's substitutionism and Svoboda's expressivism in more detail to determine whether or not these two solutions are as appealing as they appear.

VII. Critical Evaluation

In this section, I argue that revisionary moral expressivism cannot successfully avoid the Frege-Geach problem. I will then turn to substitutionism, keeping in mind the shortcomings of the solutions discussed previously, and take issue with Lutz's substitutionist solution before developing what I call hybrid substitutionism in the following section.

Before critically evaluating revisionary moral expressivism, I have to explain the Frege-Geach problem as it will play a large role in showing the implausibility of Svoboda's solution. The Frege-Geach problem illustrates the issue that non-cognitivist moral judgments have with logical inference.⁴³ To exemplify this, consider the following:

P1. If killing is bad, then killing my friend is bad

P2. Killing is bad

C. Killing my friend is bad

Using basic modus ponens, with cognitivist truth-apt language, we can see how the conclusion follows from the premises. However, this is not what the non-cognitivist picture entails. What the non-cognitive argument would look like is this:

P1. If killing is bad, then killing my friend is bad

⁴¹ Stevenson (1937)

⁴² I explain what this means in the next section.

⁴³ Geach (1965)

P2. “Boo killing!”

C. ???

The upshot is that this is not even an argument. The antecedent in the first premise is moral, but is not expressing anything because it is unasserted and the second premise has no way to confirm the antecedent in the first premise since it has no propositional content and, thus, cannot be evaluated for truth-value. This brings up the necessary questions of how the rules of logic apply to non-cognitive expressions (if, at all) and how the expressivist can account for unasserted “expressions.”⁴⁴ Without explaining how non-cognitive judgments work in conditionals, there can be no logical inference or entailment between non-cognitive judgments and, as a result, run-of-the-mill expressivism cannot even get off the ground logically speaking.⁴⁵

With this in mind, I will now evaluate revisionary moral expressivism. If one grants that non-cognitive expressions are moral judgments,⁴⁶ then this gives us prima facie reason to think it is susceptible to the Frege-Geach problem. The whole project of giving an adequate solution to the ‘Now What?’ Problem revolves around what the error theorist *should* do. If revisionary moral expressivism does not allow for logical inference or entailment, and thus a type of reasoning and justification, then how can it provide a guide for what the error theorist should do? Svoboda addresses this by arguing that, although logical inference cannot be secured, the expressivist can appeal to pragmatic consistency in one’s non-cognitive attitudes (2015, p. 22). For example, he uses the example of a person who disapproves of capital punishment (whose utterance would be something like “boo capital punishment”), but expresses satisfaction at the execution of an especially devious criminal (“hooray for capital punishment”). Svoboda argues that this would be a pragmatically inconsistent set of attitudes. Furthermore, he argues that this person’s peer could point her to the pragmatic inconsistency and they could reason as to how her attitudes *should* be.

⁴⁴ The scare quotes are because the antecedent is not really being expressed.

⁴⁵ There has been work done by Gibbard (2003) (norm-expressivism) and Blackburn (1993) (quasi-realism) addressing the Frege-Geach Problem, but my concern here, and for the rest of the paper, is with the expressivism promoted by Svoboda.

⁴⁶ Which is not uncontroversially obvious given that “boo murder” is in no way recognizably moral.

While this appeal to pragmatic consistency might seem to be a clever way to respond to the Frege-Geach problem, it does not succeed. What it actually does, I will argue, is either makes moral reasoning arbitrary or, in self-undermining fashion, implicitly posits truth-apt beliefs. Regarding the former claim, there cannot be any real, strongly justifiable consistency without some kind of logical inference or entailment. Taking Svoboda's capital punishment case, it may appear that the person who expresses disapproval at capital punishment, but expresses approval at a specific execution, is being inconsistent, but how can this be proven? Without a logic of pragmatic evaluations that accounts for non-cognitive expressions (which does not exist to my knowledge) this cannot be done. If one yells "boo" when asked about the permissibility of murder in one case this does not give any information to anyone about how this person will react to other cases (including the person uttering the expression). First, because it is a total mystery as to what the justification of the utterance is and, secondly, because there is no way to know what is entailed by such an utterance. Further, these problems compile when the expressivist wants to know what she should do. Without any kind of grounded logical inference, there is no way to know what to do next such that one would be consistent. There is not reason to think non-cognitive expressions are even consistency-apt! They do not admit of logical consistency and I do not believe there are any models that can account for such expressions. The only way to avoid this would be to posit truth-apt beliefs, but that would obviously undermine the whole project. This idea that pragmatic consistency, without logical reasoning, can serve as a framework for evaluating non-cognitive expressions is implausible.

This results in all pseudo-consistent-expression-judgments being either arbitrary or truth apt. Because there is no way to evaluate the consistency of these expressions, the evaluator of consistency is arbitrarily deciding what *seems* correct. No good solution to the 'Now What?' Problem will allow moral reasoning to be groundless. The revisionary moral expressivist must either bite the bullet that she is allowing for arbitrary

reasoning or that she is positing truth-apt beliefs that configure logically. Neither leaves the expressivist in a good position.

To summarize, the expressivist appeals to pragmatic consistency as a way to get around the Frege-Geach problem, but this leaves them in a situation where consistency evaluations are either arbitrary or positing truth-apt beliefs (and, hence, undermining). As a result, revisionary moral expressivism cannot get out of the hole that the Frege-Geach problem digs for it.

Before evaluating substitutionism, it is important to recap what is preferred in a solution to the ‘Now What?’ Problem based on the various problems previous solutions have incurred. If we take objections to the previous solutions seriously, then it is important that substitutionism does not ignore these concerns. Also, as the last remaining solution that I consider, an in-depth evaluation is necessary to see if substitutionism is the best solution and, if not, where to go next.

Svoboda evaluates his revisionary moral expressivism in terms of avoiding moral error and incurring interpersonal and intrapersonal costs. Lutz works similarly, but puts this in terms of what the error theorist should think, say, and do. Further, there are additions that should be added such as internal coherence, retaining the benefits of morality without giving up epistemic virtues, allowing for disagreement and reasoning, avoiding irrationality, providing an account of motivation, and allowing for the ability to engage in first-order moral discourse. It might seem an impossible task to get all these concerns to mesh, but it is important that the best solution take account of the advantages and disadvantages witnessed in previous solutions.

To quickly remind the reader of substitutionism’s basic tenets, Lutz argues that a good solution to the problem will not endorse any positive moral beliefs, allow the error theorist to be able to act morally and use moral language when appropriate, and action and language cannot be based on the acceptance of a moral proposition. Further, the substitutionist will develop a non-moral framework that captures the non-negotiable

and negotiable commitments of morality without the defective commitment to categorical prescription (this is called the salvaged concept). She will then become committed to this non-moral salvaged concept and use surface cognitivist grammar to refer to whatever the substituted salvaged concept encompasses. Lutz takes substitutionism to avoid moral error by using the cognitivist surface grammar to refer to a non-moral substitute. Further, the same surface grammar allows the substitutionist to engage in moral discourse and hold on to the benefits of morality (social coordination, etc.). The way substitutionism allows for disagreement, reasoning, and motivation depends on what substitutionist solution is substituted, but, at the very least, the solution offers a template for these virtues.

Looking at Lutz's criteria, they seem to provide a template that avoids the issues of previous attempts. Not endorsing positive moral beliefs avoids systematic moral error and irrationality tied to holding two contradictory beliefs. Acting morally and using moral language allows the substitutionist to engage in moral discourse and work in a "moral" framework that allows her to act "morally" without being committed to categorical prescription (thus, holding on to the benefits of moral discourse, without the defective commitment). Lastly, not taking a fictional attitude toward moral propositions allows her to avoid the irrationality of acting on the make-believe.

The first objection Lutz addresses is that the substitutionist, in using cognitivist surface grammar to refer to the salvaged concept, is being deceptive. Lutz responds to this by arguing that, because the salvaged concept that the substitutionist uses will ordinarily have commitments that overlap with the other person's, there will usually be no issue (given a lack of significant moral differences) (2014, p. 366-367). Further, this is not a level of deception that makes moral discourse impossible. If a substitutionist is asked whether or not they agree that a specific murder case is heinous, her response does not need to reveal her full philosophical views on morality. Responding in a way that makes it obvious whether there is agreement or disagreement is sufficient for first-order moral talk. Thus why the cognitivist surface grammar is retained in some respect.

The deception at this level is not an issue and if there are further concerns the substitutionist can drop her surface grammar and explain her stance.

There are other concerns that I will address in the next section, but as a template substitutionism provides adequate accounts for all the concerns provided by the dialectic. The next step is providing a specific substitute that can do everything that the template claims substitutionism is capable of. In the next section, I will consider Lutz's Humean substitutionism and, ultimately, decide against it in favor of my own flavor of substitutionism.

VIII. Hybrid Substitutionism

In this section I explain Lutz's Humean substitutionism and argue that it is not the best type of substitutionism. Next, I will put forth my hybrid substitutionism and contend that it offers the best solution to the 'Now What?' Problem.

On Lutz's account, the Humean substitutionist uses cognitivist surface grammar to assert one's approval or disapproval ("murder is wrong" is actually "I disapprove of murder") and, subsequently, is motivated to act in accordance with what she approves of. According to Lutz, this substitutionist picture provides adequate reason-giving force⁴⁷ and allows for moral reasoning based on one's desires and reactions. Because most agents have desires to act "morally," they will surely be disgusted, and disapprove of, actions such as murder, theft, etc. providing them with frameworks that do not permit generally heinous actions. Further, first-order disagreement can be accounted for via a contention in attitudes.⁴⁸ This version of substitutionism avoids moral error and irrationality, and allows the substitutionist to engage in moral discourse and be motivated to act in moral ways, but it comes up short in offering an adequate account of moral reasoning, disagreement, and action, which does not allow for heinous activity.

⁴⁷ As illustrated earlier, using approval and disapproval provide motivation in the same way one's approval or disapproval of certain tastes motivates them to eat certain foods.

⁴⁸ Stevenson (1937)

Basing action solely on one's approval or disapproval, and presupposing, as Humean substitutionists do that all agents hold similar attitudes towards various human behaviors, does not guarantee "moral" action. Even if an agent desires to act morally, this does not entail that she will disapprove of all heinous actions. Further, attitudes of approval and disapproval are not necessarily grounded in something concrete. One's disapproval of murder could simply be based on murder feeling *icky* to them. This is not a concrete enough way to prevent murderous motivation and similar motives. Also, this kind of justification seems very flexible based on one's situation. While it might hold up most of the time, it is not hard to imagine that when one is betrayed or in an angry outrage, murder might not seem so icky. Thus, given the loss of disapproval of the action, an agent might not have any real reason to not murder. It seems there should be something more to protect against this kind of allowance. Further, this kind of icky-reasoning might allow for disagreement in the most shallow sense, but when two people are disagreeing and one argues that a certain action is icky and the other that it is not, how do they reach resolution? Icky-ness is not something that requires logical entailment; it is simply a reaction people have. I cannot imagine one can be convinced that something is not icky if she has icky reactions to it. Lastly, for a framework of moral reasoning, this kind of allowance for icky-reasoning is weak. There should be something more than, "this feels weird to me" to ground action, reasoning, and disagreement.

As a result, I will argue for what I will call "hybrid substitutionism." In a nutshell, hybrid substitutionism takes the attitudes introduced by Lutz (of approval and disapproval) and supplements them with a type of practical rationality that accounts for one's short, middle, and long-term goals. To illustrate the advantages of this position, it is helpful to consider the challenge posed by the case of Hume's sensible knave. The sensible knave is a person who, in the context of contractarianism, has the opportunity to defect from the social contract in order to benefit himself (at the risk of being punished by the sovereign).⁴⁹ The

⁴⁹ In this case, the sovereign is the central power that penalizes defections. An easy way to think about this is to substitute in the modern justice system for the sovereign.

challenge, as applied to the ‘Now What?’ Problem, is to explain why the knave should not go through with their knavery (theft, for example). Lutz’s substitutionism would hold that, because most agents disapprove of such actions, stealing for his own good would disgust the knave. However, this is where Lutz’s substitutionism comes short. What if the knave is not disgusted and holds no disapproval-like attitudes towards theft? This form of substitutionism has no response and, thus, cannot adequately respond to the challenge posed by the sensible knave. Hybrid substitutionism, on the other hand, can. Hume’s response to the sensible knave is that it would be irrational for the knave to defect. The knave, if rational, will realize that defecting is too great of a risk to his goals, which likely include preserving one’s reputation, avoiding imprisonment and maintaining peace of mind.⁵¹ It is this kind of rationality that hybrid substitutionism adopts and uses to address the challenge of Hume’s sensible knave.⁵²

Applying this consideration of the knave to the ‘Now What?’ Problem, it seems that supplementing the approval and disapproval attitudes with the type of practical rationality illustrated by the rational sensible knave allows for the hybrid substitutionist to offer adequate answers to all the concerns raised by the ‘Now What?’ Problem. While one might not feel disgust or icky-ness in reaction to the action, or thought, of murder, their knowledge of the consequences that come with murder will be enough to prevent them from acting in such a way. This might seem like a weak solution since it might imply that the hybrid substitutionist is a possible murderer barely held back by their practical rationality, but the hybrid substitutionist has two ready responses. First, this only applies to those who are not disgusted by murder (which is already a small amount) and, second, preventing murder is preferable to not doing so.

The hybrid substitutionist takes Hume’s response to the sensible knave’s challenge seriously and will introduce a kind of practical rationality into her substitutionist framework. She will figure out what is in her

⁵⁰ To make things clearer, a contemporary version of the sensible knave would be someone who knows they have a good, but not assured, chance at getting away with breaking a law (such as stealing) that results in their benefit (possibly monetarily).

⁵¹ Hume (1777)

⁵² Furthermore, both Joyce (2001) and Garner (2007) argue that Hume’s response to the sensible knave is “roughly correct” and that the knave has “good instrumental reasons for acting in a cooperative manner” (p. 210)

short, middle, and long term goals and desires and will act in accordance to what furthers these with consideration to her approval and disapproval attitudes and also what is rational for her. Her ability to succeed and flourish “morally” will be supported by her substitutionist framework, which allows her all the benefits of first-order discourse without any of the defective commitments that error theorists take issue with. The hybrid substitutionist avoids moral error by using cognitivist surface grammar to refer to a combination of non-moral approval or disapproval in supplement with the non-moral practical rationality (“not only am I disgusted by murder, but it also is not rational for me to do so”). It allows the substitutionist to engage in first-order moral discourse via that cognitivist surface grammar and avoids any kind of awkwardness while holding onto the benefits of moral discourse. Further, it allows for motivation via approval and disapproval attitudes along with knowledge of what is rational to do. Disagreement is accounted for by contending attitudes and justification can be traced not only to what one is disgusted by, but also by what is rational for one to do (which is much more workable when in contention than some feeling of icky-ness!). There is no hint of irrationality or internal inconsistency and, lastly, there are no concerns of psychological impossibility. All things considered, hybrid substitutionism is the best answer to the ‘Now What?’ problem.

IX. Conclusion

In summary, I’ve explained the ‘Now What?’ Problem, given the dialectical background, argued against revisionary moral expressivism and Humean substitutionism, and developed hybrid substitutionism as the best solution to the problem. This solution should allow the error theorist to “fill the hole in their normative life”⁵³ and make moral error theory’s consequences seem much less dramatic than they are usually taken to be.

⁵³ Lutz (2014, p. 370)

Bibliography

- Blackburn, Simon (1993). *Essays in Quasi-Realism*. Oxford University Press.
- Brandt, Richard B. (1998). *A Theory of the Good and the Right*. Prometheus Books.
- Finlay, Stephen (2008). The error in the error theory. *Australasian Journal of Philosophy* 86 (3):347-369.
- Finlay, Stephen (2011). Errors upon errors: a reply to joyce. *Australasian Journal of Philosophy* 89 (3):535 - 547.
- Foot, Philippa (1972). Morality as a system of hypothetical imperatives. *Philosophical Review* 81 (3):305-316.
- Garner, Richard (2007). Abolishing Morality. *Ethical Theory and Moral Practice* 10 (5):499-513.
- Geach, P. T. (1965). Assertion. *Philosophical Review* 74 (4):449-465.
- Gibbard, Allan (2003). *Thinking How to Live*. Harvard University Press.
- Hinckfuss, I. (1987). *The moral society: its structure and effects (Vol.16)*. Canberra: Australian National University, Departments of Philosophy.
- Hume, David (1777). *An Enquiry Concerning the Principles of Morals*. Prometheus Books.
- Jaquet, François & Naar, Hichem (2016). Moral Beliefs for the Error Theorist? *Ethical Theory and Moral Practice* 19:193-207.
- Joyce, Richard (2001). *The Myth of Morality*. Cambridge University Press.
- Joyce, Richard (2011). The Error In 'The Error In The Error Theory'. *Australasian Journal of Philosophy* 89 (3):519-534.
- Joyce, Richard (2012). Enough with the errors! A final reply to Finlay. Unpublished. 1-15.
- Kant, Immanuel (1785/2002). *Groundwork for the Metaphysics of Morals*. Oxford University Press.
- Köhler, Sebastian & Ridge, Michael (2013). Revolutionary Expressivism. *Ratio* 26 (4):428-449.
- Lutz, Matt (2014). The 'Now What' Problem for error theory. *Philosophical Studies* 171 (2):351-371.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Penguin Books.

Nolan, Daniel ; Restall, Greg & West, Caroline (2005). Moral fictionalism versus the rest. *Australasian Journal of Philosophy* 83 (3):307 – 330.

Olson, Jonas (2011). Getting Real about Moral Fictionalism. *Oxford Studies in Metaethics: Volume 6* 6:181.

Olson, Jonas (2014). *Moral Error Theory: History, Critique, Defence*. Oxford University Press.

Stevenson, Charles Leslie (1937). The emotive meaning of ethical terms. *Mind* 46 (181):14-31.

Streumer, Bart (2013). Can We Believe the Error Theory? *Journal of Philosophy* 110 (4):194-212.

Svoboda, Toby (2015). Why Moral Error Theorists Should Become Revisionary Moral Expressivists. *Journal of Moral Philosophy*:1-25.



