

# The Yale *P*HILOSOPHY REVIEW

*An Undergraduate Publication*

ISSUE 6 | 2010

---

**Propositions, Clarification, and  
Faultless Disagreement**

BRENDAN DILL, *Yale University*

**Dispositional Properties and  
Humean Supervenience**

JEREMY GOODMAN, *Brown University*

**Bringing a Text to Life: The Role of the  
Reader in Plato's *Phaedrus***

MELISSA TAN, *University of Chicago*

**The Third Man Argument: *Parmenides* 132a1-b2**

IAN WELLS, *Cornell University*

**Interview with Daniel Dennett,  
Tufts University**

---

## §

The *Yale Philosophy Review* is an annual journal that showcases original philosophical thought by undergraduate students, worldwide. The goal of the *Review* is to promote philosophical discourse of the highest standard and to bring together a community of young philosophers both in the United States and abroad. Each issue contains a selection of essays on a broad range of topics as well as an interview with a notable contemporary philosopher.

---

ISSUE VI, 2010

**EDITORS-IN-CHIEF**

Chandler Coggins  
Hayley Johnson  
Geoffrey Shaw

**EXECUTIVE EDITORS**

Jacob Abolafia  
Joseph Breen  
Jeremy Lent  
Susan Morrow  
Justin Petrillo

**COPY EDITORS**

Tobias Kuehne  
Jeremy Lent  
John I. Robinson  
Camila Yadeau

**EDITORIAL BOARD**

Annie Atura  
Ilan Ben-Meir  
Ronik Bhargoo  
Katie Carmody  
Alex Cohen  
Johnny Copp  
Andrew Eberle  
Harris Eppsteiner  
Eitan Fischer  
Noah Gentele  
Simon Goldstein  
Sisira Gorthala  
Stephen Grugett  
Charles Gyer

Karen Huang  
Samuel Jacobson  
Geoffrey Liu  
Elliot Milco  
Jack Montgomery  
Christian Mott  
Natasha Paterson  
Lucas Pratt  
Sudhir Rao  
Jeremy Shifferes  
Maxwell Silva  
Adrienne Spiegel  
Matthew Wagstaffe  
Camila Yadeau

**FACULTY ADVISORS**

Michael Della Rocca  
Shelly Kagan

For online archive and submissions guidelines, visit <[www.yale.edu/ypr](http://www.yale.edu/ypr)>.  
Email the editors at [yalephilosophy@gmail.com](mailto:yalephilosophy@gmail.com).

---

## EDITORS' NOTE

The *Yale Philosophy Review* is now in its sixth year of life. Founded by a group of Yale undergraduates, the *Review* continues in its mission to constitute an active community of college students doing philosophy with one another.

In addition to producing this annual publication, we organize a yearly Yale College student essay contest and host Yale professors for campus-wide discussions of various philosophical topics and questions.

We are grateful to have received over one hundred submissions to this year's *Review*, from students the world over. Over thirty-five undergraduate board members carefully read and weighed the merits of every paper in five different committees organized by subject matter. We approached the discussions with seriousness and rigor, and we quite enjoyed them, too. Turns out there's a lot of good philosophy being done by college students.

We trust this issue of the YPR will challenge and excite you as much as it has us. The four essays included here cover topics in the history of philosophy, philosophy of physics, and philosophy of language. Finally, our interview with Daniel Dennett treats, among other things, philosophy on sailboats.

As always, we thank the Yale Philosophy Department for its support.

Chandler Coggins, Hayley Johnson, & Geoffrey Shaw  
Editors-in-Chief  
*The Yale Philosophy Review*

---

# The Yale Philosophy Review

Issue VI, 2010



## CONTENTS

- 5        **Propositions, Clarification, and Faultless Disagreement**  
          BRENDAN DILL, YALE UNIVERSITY
- 34       **Dispositional Properties and Humean Supervenience**  
          JEREMY GOODMAN, BROWN UNIVERSITY
- 46       **Bringing a Text to Life: The Role of the Reader in**  
          **Plato's *Phaedrus***  
          MELISSA TAN, UNIVERSITY OF CHICAGO
- 56       **The Third Man Argument, *Parmenides* 132a1-b2**  
          IAN WELLS, CORNELL UNIVERSITY
- 88       **Interview with Daniel Dennett, Tufts University**

---

# Propositions, Clarification, and Faultless Disagreement

BRENDAN DILL  
*Yale University*

Abstract: Both contextualist and relativist solutions to the faultless disagreement problem clash with our intuitions: contextualism, with the intuition that two people arguing about a matter of taste are in fact disagreeing; and relativism, with the intuition that the truth of a proposition is independent of who is evaluating it. In this paper, I will outline a solution that explains our intuition of disagreement without clashing with our intuitions about truth. I will do this by proposing a definition of propositions as ideally clarified assertoric content, having one absolute truth-value that does not vary across any contexts. I will argue that this definition is plausible, that it best serves the purposes of philosophy, and that it best solves the problem of faultless disagreement.

## I. Faultless Disagreement, Contextualism, and Relativism

As posed by Crispin Wright, the problem is this: John and Mary are at dinner together and both see that stewed rhubarb is offered on the menu. Excited, John says, “rhubarb is delicious!” Mary makes a face and responds, “rhubarb is *not* delicious.”<sup>1</sup>

Wright says that we have a threefold intuition about such a dispute, which he calls “the Ordinary View.” First, we believe that John and Mary’s attitudes are “genuinely incompatible,” that they contradict each other, or, that they disagree. Second, we believe that neither John nor Mary “need be mistaken or otherwise at fault,” in other words, that their disagreement is faultless in nature. This is because of the apparent subjectivity of the judgment that rhubarb is delicious—most of us are comfortable with the fact that taste varies from person to person. The third intuition is that both positions are rationally sustainable, that neither John nor Mary need withdraw their assertions in light of the disagreement.<sup>2</sup> The first two intuitions, of contradiction and faultlessness, are sufficient to pose the problem that contextualism and relativism attempt to solve.

---

<sup>1</sup> Wright 2006, p.38.

<sup>2</sup> Wright 2006, p.38.

---

There is a simple argument that seems to show that contradiction and faultlessness cannot both be true: if rhubarb is delicious, then Mary is at fault; if she is not at fault, then rhubarb must be not delicious, and thus John must be at fault. In general, this seems to show that if Mary asserts a proposition P and John asserts Not-P, then one of them must, logically, be wrong.<sup>3</sup>

There are many solutions to this problem, as mapped out by Max Kölbel in “Faultless Disagreement.”<sup>4</sup> Some use intuitionistic logic; others take statements like “rhubarb is delicious” in an expressivist fashion, saying that they are not truth-apt but merely prescriptive. Others stick to the realist position that there is an absolute fact of the matter about deliciousness, and so either John or Mary is wrong. I will ignore these, as others have given ample arguments showing their shortcomings, and focus on contextualism and relativism as the two best current contenders to solve this problem.

Contextualism throws out disagreement in favor of faultlessness, saying that in this case John and Mary’s assertions have been misrepresented. When John says “rhubarb is delicious,” his meaning is affected by his own context, that is, his standard of taste, so that he really *means* “rhubarb is delicious *to me*.” Contextualism takes various forms, in which “rhubarb is delicious” can mean “rhubarb is delicious by the standards of taste of my group” or “rhubarb is delicious by the standards of taste of most experts” as well as the simpler “rhubarb is delicious to me.”<sup>5</sup> In any of these cases, the *content* of the proposition that John asserts is dependent on John’s own context in a systematic way, either by a hidden indexical or by context-dependent changes in the meaning of the word “delicious.”

However it is formulated, however, contextualism runs against the same objection. There is a strong intuition that these two disputes are fundamentally different:

- (1) John: Rhubarb is delicious.  
Mary: Rhubarb is not delicious.
- (2) John: Rhubarb is delicious to me.  
Mary: Rhubarb is not delicious to me.

---

<sup>3</sup> See Wright 2006 and Kölbel 2003 for more formal versions of this proof.

<sup>4</sup> Kölbel 2003.

<sup>5</sup> Kölbel 2003, p. 63.

---

Dispute (2) simply does not seem to be a dispute at all! As John MacFarlane notes, “if in saying ‘apples are delicious’ I am saying that they taste good *to me*, while in saying ‘apples are not delicious’ you are denying that they taste good *to you*, then we are no more disagreeing with each other than we would be if I were to say ‘My name is John’ and you were to say ‘My name is not John.’”<sup>6</sup> Even if it is a mistake about the meanings of our words, the contextualist needs to account for this apparent difference in disputes if she is going to defend her theory from the charge of gravely misrepresenting linguistic practice. So far, it seems that the contextualist has no way to provide this defense and remains unable to account for people arguing with each other about assertions which do not mutually conflict.<sup>7</sup>

It is this inadequacy in the contextualist solution that the relativist aims to answer. The truth-relativist or propositional relativist (to distinguish from other views called “relativism”) believes that the truth of some propositions varies relative to a context of assessment. In other words, the very same proposition *P* can be true *for me* and false *for you*. Thus he provides a solution to the faultless disagreement problem that meshes with our intuitions: both Mary and John are disagreeing over *the very same proposition* (John asserting *P*, Mary asserting Not-*P*), and, since the proposition’s truth is relative to their own context of assessment, they are *both* asserting something true. They are genuinely contradicting one another, but neither is at fault. Problem solved.

The relativist argues that the truth of propositions is already relative to some parameters: possible worlds under most views, and under some views, time and epistemic standards. MacFarlane calls these parameters a “circumstance of evaluation,” and says that it is “standard practice” to relativize propositional truth to these.<sup>8</sup> The relativist merely takes this one step further, relativizing truth to a parameter that shifts with “the context in which the speech act (or other use of the sentence) is being assessed.” MacFarlane continues:

In order to state the relativist’s position, then, we must employ the doubly contextual predicate ‘true at context of use  $C_U$  and context of assessment  $C_A$ ’ in place of the familiar ‘true at context

---

<sup>6</sup> MacFarlane 2007, p. 18.

<sup>7</sup> One contextualist account that Kölbel mentions is the idea of a conversational scoreboard (put forward by David Lewis) to which these disputes are relative to, but Kölbel observes rightly that we can think of people disagreeing who aren’t directly conversing and thus don’t share a conversational scoreboard.

<sup>8</sup> MacFarlane 2005, p.323.



---

of use  $C$ '. By a 'context of assessment', I mean simply a concrete situation in which a use of a sentence is being *assessed*.<sup>9</sup>

Since standards of taste are the type of thing that varies with contexts of assessment, John's assertion "rhubarb is delicious" can be true at his context of assessment, and Mary's assertion "rhubarb is not delicious" can be equally true at hers. This variation of truth is not only true of the sentences they say (something which is provided by contextualism), but of the *propositions* those sentences express.

The challenge, as MacFarlane notes, is to give an account of what it means to be "true at context of use  $C_U$  and context of assessment  $C_A$ ." Rather than looking for an account of this via a definition of truth in more primitive terms, MacFarlane argues that our best conception of truth might come from looking at the commitments of assertion: "an assertion (even an insincere one) is a *commitment to the truth* of the proposition asserted."<sup>10</sup> MacFarlane proposes that this commitment is composed of three sub-commitments: to withdraw the assertion if it is shown to be false, to justify the truth of the assertion in the face of putative challenges, and to be held responsible if another takes action based on the assertion and that assertion turns out to have been false.<sup>11</sup> These take relativist forms by substituting "true" or "false" for "true/false relative to the context of use and context of assessment." Thus the commitments of an assertion encompass other contexts in which the proposition may be assessed, either by the asserter or by others. This makes sense: if one is asked to justify or withdraw an assertion they made in another context, they must justify it or be shown its falsity relative to their current context of assessment, rather than simply responding "it was true *in that context*." Thus MacFarlane argues that the doubly-contextual truth-predicate of the relativist has been made intelligible, and even plausible, in light of our norms of assertion.

In order to solve the problem of faultless disagreement, however, the relativist view requires a new view of disagreement. The conventional view of disagreement is simply that two people disagree just in case there is a proposition  $P$  such that one asserts  $P$  and the other asserts Not- $P$  (or, in other terms, such that one accepts  $P$  and the other rejects  $P$ ). MacFarlane argues that this conception

---

<sup>9</sup> MacFarlane 2005, p.325.

<sup>10</sup> MacFarlane 2005, p.333.

<sup>11</sup> MacFarlane 2005, p.334.

---

cannot be true, if propositions are relative to time or worlds, for the proposition that Joe is sitting now can be asserted at 2 PM and denied at 3 PM without any real disagreement going on; or similarly, the proposition that Mars has two moons can be asserted in one possible world while denied in another possible world without disagreement going on either. Using norms of assertion once again, MacFarlane presents a relativist view of disagreement. He says that to assert something is to commit oneself to its “accuracy.” Accuracy is a relative property—an acceptance or rejection of a proposition is accurate iff that proposition is true or false relative to the context of use (i.e., the world in which it is asserted) and the context of assessment (i.e., the standard of taste of the assessor). Thus, at any one context of assessment, both Mary and John’s assertions cannot both be accurate, though they can both be asserting something true relative to their own contexts of assessment. This gives a plausible picture of faultless disagreement:

The challenger thinks (rightly) that he has absolutely compelling grounds for thinking that the assertion was not accurate. But the original asserter thinks (also rightly, from her point of view) that the challenger’s grounds do nothing to call in question the accuracy of the assertion. The asserter’s vindication will seem to the challenger not to show that the assertion was accurate, and the challenger will continue to press his claim. (Until the game gets boring.) Thus we have all the normative trappings of real disagreement, but without the possibility of resolution except by a relevant change in one or both parties’ contexts of assessment.<sup>12</sup>

Thus MacFarlane has given a plausible relativist picture of truth and disagreement, which, unlike contextualism, allows for Wright’s Ordinary View to hold true.

Though relativism has been shown to be plausible and internally consistent, one must consider the philosophical costs of adopting such a position. I believe that in order to save our intuition that Mary and John are having a faultless disagreement, relativism sacrifices stronger intuitions about the nature of truth, disagreement, and propositions.

---

<sup>12</sup> MacFarlane 2007, p.29.

---

Though some may look down on this as a naïve view, many (including myself) have the strong intuition that truth is something absolute and objective. Adding in dependence on contexts of assessment seems to make truth remarkably subjective. Nowhere in the relativist picture do I see an account of the difference between what it is to be *true for A* and what it is to *seem true to A*. Maybe for questions like whether rhubarb is delicious the two are equivalent—if rhubarb seems delicious to John, then it *is* delicious from his context of assessment. But what is to prevent this truth-predicate from spreading to areas in which it is a grave misrepresentation of assertoric practice? For instance, if someone asserts “concentration camps were bad,” they will not think it legitimate to say that this assertion is not true from the context of assessment of Hitler. The relativist response is to apply the doubly-indexed truth predicate in all domains, but say that in some domains, like personal taste, the truth of a proposition varies with the context of assessment, while in other domains, like morality or science or at least mathematics, the truth of a proposition does not vary with the context of assessment. But the question then is how to distinguish between domains in which truth-values do and do not vary with contexts of assessment. In other words, what is the difference between the truth of a proposition varying with context and opinions about the truth of a proposition varying with context? This is the same question raised at the beginning of this paragraph—and while the relativist may have a semantic way of dealing with a distinction between the two, there seems to be no non-arbitrary way of supporting that distinction. The mere presence of intractable disagreements in a particular domain is not sufficient, I believe, for this distinction. So, until a relativist can give an account of this distinction, the doubly-indexed truth-predicate seems worrisome for those who wish to preserve the objectivity of truth.

It also seems impossible to state the relativized truth-predicate without reference to our conventional absolute truth-predicate. In accordance with a relativist picture, in what sense are meta-context claims, like the proposition that neither Mary nor John are mistaken, *true*? If the truth of the proposition “rhubarb is delicious” is always relative to a particular context of assessment, then at each context of assessment the proposition will either be true or false; therefore there is no context of assessment at which one can truthfully say that neither Mary nor John are mistaken. As Wright points out, in fact, if Mary and John genuinely disagree, then they must each think that the other is mistaken, and so anyone who regards the disagreement as faultless must in fact regard all parties involved as mistaken, for each has attributed a fault to the other that does

---

not exist.<sup>13</sup> In order to regard neither side as mistaken, then, it seems that one needs to use an assessment-independent truth predicate to describe the assessment-dependent truth of each assertion. The outside listener, no matter his personal opinions on rhubarb, will be able to say that P is *true for* John and Not-P is *true for* Mary (in the assessment-dependent sense), and that these facts about John and Mary's relations to the proposition P are *true* (in the absolute sense). This need for the absolute truth predicate in some situations necessitates a stronger relativist account for how to distinguish between situations in which "true" is a doubly-contextual predicate and situations in which "true" is an absolute predicate.

The accept/reject definition of disagreement that MacFarlane throws out so easily is an intuition that deserves to be taken more seriously. When two people are disagreeing, they think of it in terms of there being a common ground that they are contesting over—namely, a proposition with one set of truth-conditions, which one person accepts and the other rejects. There is a strong intuition that there is a perspective-independent fact of the matter that is being contested. I do not see how MacFarlane's view of disagreement as contested perspectival accuracy, an irresolvable game, is any farther from the idea that John and Mary are "talking past one another" than contextualism. Though they may be disagreeing over the same proposition, it is affected by their respective contexts of assessment to the extent that what John is accepting does not have the same truth-conditions as what Mary is rejecting, hence why both of their assertions can be "true." It seems that if we are to call this disagreement, then we need to change our notion of disagreement drastically.

This leads into the final worry: that relativism clashes with intuitions about what propositions should be. The intuition, simply stated, is that if the same proposition is true in one context of assessment, and its negation is true in another context of assessment, then in fact it is not one, but *two* separate, albeit closely related, propositions. The relativist definition of propositions does not provide for this. As far as I can see, MacFarlane defines propositions as the contents of mental states such as beliefs and desires. This allows for a distinction between propositions and utterances—propositions are content, while assertoric utterances are the form in which they are expressed. The problem is that this definition of propositions does not specify how specific the content of a belief must be, and thus allows the content or truth-conditions of beliefs to be as ambiguous as the utterances which express them. Another way of putting this is

---

<sup>13</sup> Wright 2006, p.52.

---

that the line between what parameters we can and cannot make the meanings of propositions relative to seems arbitrarily drawn. There seems to be nothing to keep a relativist from saying that “John went to the bank today” expresses one proposition, that John went to the bank today, the truth-value of which varies depending on whether the speaker is speaking about John’s visit to the side of a river or to a financial building. The immediate objection is that “John went to the bank today” really expresses one of two propositions, depending on the meaning of “bank.” But those two propositions are distinct not because of the syntactic structures of the sentences expressing them, but because they have different truth-conditions. The relativist definition of propositions as the content of beliefs does not specify that these beliefs must have one set of truth-conditions, and thus needs not distinguish between these two propositions. But it feels like a distinction between the two is needed even for the content of a belief. If the content of beliefs has to do with what is true, then how do we specify what exactly a person believes is true without one unambiguous set of truth-conditions? So it seems that even with propositions as the content of belief, we want to distinguish between two propositions that have different truth-conditions, though they may be expressed by the same utterance. But if two propositions are distinct because of their different truth-conditions, then it seems we must also say that “rhubarb is delicious” *also* expresses two propositions, depending on whether it is assessed from John or Mary’s perspective. The distinction between utterances and propositions thus seems to lack a necessary dimension, if both are allowed to be ambiguous between two different truth-conditional meanings.

Another way of stating this worry is the intuition that “propositional truth depends exclusively on the properties of and relations among the entities the proposition is about.”<sup>14</sup> If we believe this, which seems a reasonable standard for propositions, then a relativist view of propositions seems even stranger. From the relativist perspective, the relation between rhubarb and the predicate “delicious” is not enough to determine the truth-value of the proposition that rhubarb is delicious. Rather, the truth value of the proposition varies dependent on the standard of the taste of the person assessing it. The standard of taste of the assessor is not a property of or relation among the entities the proposition is about—rather, it is a feature of a truth-predicate we have already seen runs counter to intuitions about truth. It seems we could save our intuitions about truth, disagreement, and propositions all by following this intuition and making anything which the truth of a proposition is dependent upon an explicit part of the

---

<sup>14</sup> Taken from Zoltan Szabó, handout in class.

---

content of that proposition. In other words, by bringing contexts of use and assessment out of hiding in the truth-predicate and into the assertoric content of propositions, we can keep our simple notion of truth as well as a clearer distinction between propositions and utterances, and thus have a better framework with which to view disagreement.

I propose that this is exactly what we should do.

## II. Propositions as Ideally Clarified Assertoric Content

Propositions are a philosophical tool. Sentences are murky linguistic creatures; it is hard to say of a string of words that it is true or false, and harder if that string of words could mean several different possible things. Language is also very rich, and has many ways of expressing the same thing—thus we say that when two sentences assert the same thing in different ways or even different languages, they are expressing the same *proposition*. It is for these and many other reasons that it is useful to speak not of sentences, but of the propositions they express.

I emphasize that propositions are a useful tool because they are useful for a *purpose*. It seems, then, that the best definition of propositions would be one that best serves the primary purpose for which we use propositions. So, we must ask: what are propositions a tool *for*?

This question probably has many answers. Indeed, propositions are a tool for philosophy, and so to narrow down their ultimate end would be to say that there is a single unified purpose of philosophy, a claim which I have no wish to make. But there are some central purposes that we can use to narrow down a methodological purpose for propositions that perhaps will best serve all viable purposes of philosophy.

Perhaps this can be seen in why we find sentences<sup>15</sup> to be an inadequate tool when we are doing philosophy. I believe this is for two reasons. First, talking about a sentence is not necessarily talking about its content. Since it is the *content* of “snow is white” we are interested in (when we ask whether it is true that snow is white), not the fact that it has eleven letters, and not the fact that it’s written in English, we talk about not the sentence “snow is white,” but its content, the proposition *that* snow is white. Thus propositions are a tool for

---

<sup>15</sup> Meaning sentences that express assertions. When one utters a non-assertoric sentence, like a question or command, one is obviously not asserting a proposition. So the question is why we find assertoric sentences inadequate when we are doing philosophy, and I assume propositions to be a tool for understanding assertoric content.

---

talking directly about the content of assertions, rather than the sentences that express that content. As has already been mentioned, the relativist view of propositions as the content of beliefs rectifies this first inadequacy. However, it does not address the second inadequacy of sentences, which is that sentences are often ambiguous in what content they express. One cannot evaluate the truth-value of “John went to the bank today” because it expresses two different possible propositions with different truth-conditions: either that John went to a river bank, or that John went to a money bank. Thus one can split an ambiguous sentence into two different clarified propositional contents, making propositions a useful tool for clarifying assertions.

So we’ve distinguished two features of propositions that make them useful and distinguish them from sentences or utterances: first, that propositions are the content of assertions rather than the words with which that content is expressed; and second, that the content of propositions is not ambiguous. These two features seem to gel well with two major candidates for the purpose of philosophy: first, the search for truth; second, the project of clarifying our concepts and assertions for the purposes of understanding the questions we argue about. If our purpose is to see what is true, then first we must have something to which the predicate “true” can be applied to—assertoric content—and second we must be able to narrow down the truth-conditions such that we can evaluate whether that content is true or false (hence why propositional content needs to be unambiguous). If our purpose is to clarify assertions, then first we must distinguish the form of the assertions from their content—hence the sentence/proposition distinction—and then our very project is to find one singular unambiguous content for that assertion, which we call the proposition.

With these observations I venture to give the following definition: *a proposition is the content of an assertion, which has one and only one possible truth-conditional interpretation and thus one and only one absolute truth-value.* In shorter terms, propositions are ideally clarified assertoric content.

What I mean by “ideally clarified” is that propositions are created from sentences by the process we went through to turn the sentence “John went to the bank today” into two possible propositions with different truth-conditions. I envision this process as being something like asking someone who has made an assertion, “what do you mean?”, perhaps even presenting two different interpretations that you wish for the asserter to distinguish between. This process has to be gone through until there is only one possible assertoric content, which has fixed truth-conditions which can be assessed from any perspective to yield the same truth-value. This clarification process need not be explicit, however,

---

and most or all of the work may be done by the context. Critically, however, the context does not change the truth-value of a set proposition but rather acts in translating from an utterance to the content of a proposition. Also, context may not always be sufficient to pick out one proposition, in which case the “what do you mean” process of clarification must become explicit.

One more fairly uncontroversial example: “I went to the store” does not express the same propositional content (that I went to the store) with each utterance. Rather, we find its propositional content by replacing the indexical with its descriptive meaning: if John says “I went to the store,” he asserts that John went to the store; if Mary says “I went to the store,” she asserts not that John went to the store, but that Mary went to the store. Thus the sentence “I went to the store” on its own is ambiguous<sup>16</sup> between several different assertoric contents with different truth-conditions. The knowledge of who uttered it (what MacFarlane might call the context of use) clarifies the sentence into a single proposition with one truth-value. And, just as it is impossible to evaluate the truth of “John went to the bank today” without asking whether it means that John visited a river or a financial building, it is impossible to evaluate the truth of “I went to the store” without asking who uttered the sentence. Thus the question “what do you mean” is not only a legitimate question, but a *necessary* one in order to evaluate the truth of the assertion.

This also shows us an important feature of the translation from sentences into propositions. The context of use, namely that John was the one who uttered the sentence “I went to the store,” is outside of the content of the sentence but is an explicit part of the content of the proposition. This goes back to our intuition that propositional truth depends on the things the proposition is about. It seems that an essential part of the process of translating an assertoric sentence into its propositional content is taking contextual information and making it an explicit part of the content. This is not limited to contextualism—contextualist or not, we do this on a daily basis. When asked what John asserted when he said, “I went to

---

<sup>16</sup> This is a non-standard use of “ambiguous,” which is often meant to be distinguished from indexicality, in that ambiguous sentences have a finite class of possible truth-conditional values, while indexical sentences have an open-ended, infinite class of possible truth-conditional values. What matters in terms of my discussion is simply whether a particular assertoric sentence expresses more than one set of truth-conditions. So I here and throughout the paper use “ambiguous” as a blanket term (since no other term is handy) describing anything which can express more than one semantic value or proposition, whether it is an infinite class, a finite class, or even two possible values that can be expressed.



---

the store,” we say, “he said that he went to the store.” This translation into indirect speech has implicit in it the translation of contextual information into explicit content: he said that *he* went to the store.

This clarification process is not one of making further commitments, but rather clarifying what commitments one has made. Someone might worry that the question “what do you mean?” is a legitimate one, but not a disambiguating one. Think of the following dialogue:

John: A certain guy in the office is a pain in the neck.

Mary: Do you mean Richard, Paul, or Frank?<sup>17</sup>

Mary’s question is legitimate, but it doesn’t seem like it is disambiguating between possible propositions that John might be asserting, namely that Richard is a pain in the neck, that Paul is a pain in the neck, or that Frank is a pain in the neck. For it is perfectly possible that John could be making this assertion without knowing the answer to Mary’s question—for instance, if someone spilled coffee on the copier but didn’t clean it up, and John doesn’t know who did it, but thinks that whoever did it is a pain in the neck. But this merely shows that Mary’s question is not disambiguating because it does not take all the possible propositions expressed by John’s sentence into account, by presuming that John knows which man he is accusing of being a pain in the neck. We still, however, have an ambiguous assertion on our hands, because there are several propositions it could express with different truth-conditions. In the copier case, the proposition is true iff there is a guy in the office who spilled coffee on the copier, and thus is a pain in the neck (thus it would be false if it were someone visiting the office who was the culprit). However, if John means a particular man when he says “a certain guy” (perhaps he saw Frank spill coffee on the copier), then Mary *does* need to know which particular man he is referring to in his assertion in order to evaluate its truth. For if Mary nods and says, “yeah, Richard *is* a pain in the neck,” and John meant Frank, then John will say (regardless of his opinion on Richard) “no, I meant Frank!” Mary’s interpretation was legitimate, but wrong—meaning that John’s assertion was ambiguous. This example serves to illustrate that the process of clarifying from a sentence to a proposition is not one of making *further* commitments, but rather of clarifying which of several possible commitments the asserter did and did not make.

---

<sup>17</sup> This objection was suggested to me by Zoltan Szabó.

---

The other essential observation to make about this process is that there is no reason to stop this process of moving contextual information which is outside of the sentence into the explicit content of the proposition. Any stopping point, leaving some context outside of the content of the proposition, would be arbitrary. And if our goal is to have a truth-evaluable content, then this process of clarification by making context an explicit part of content should not stop until the proposition's truth value holds independent of context, and it is no longer possible to clarify the content further.

Thus anyone who wishes to stop this clarification process at a certain point—for I will carry it into more controversial areas—must give a reason for their stopping point which does not simply appeal to their accepted definition of propositions. A definition of propositions that draws a line for clarification must show that this line is not arbitrary by appealing to a purpose that propositions serve. I have here argued that the definition that best serves the purpose of propositions puts no limit on this process of clarification—that in fact this process *must* be without limit in order to produce a truth-evaluable content. An argument against this position must not simply proceed from an accepted definition of propositions but justify that very definition with reference to its purpose—for propositions are a tool, used for a purpose. I can see three acceptable paths that can be taken against this view: first, that propositions are not a tool created by philosophers, but have an independent, objective existence, and their independent nature is not that which I have described (though how one intuits their objective nature, I do not know); second, that the purpose I have described for propositions is not the best purpose for propositions as a tool and thus my definition is at fault; and third, that the definition I have described does not best serve the purposes I outlined. Mere appeal to an arbitrary conventional definition, I believe, will not suffice.

Thus I have laid out my view of what propositions are, and how we can translate from assertoric sentences to the content that they express. In the next section, I will argue that this definition of propositions gives us a solution to the problem of faultless disagreement that is different from both contextualism and relativism, and does a better job than either of explaining some intuitions without sacrificing others. Finally, I will then examine whether it is actually *possible* to clarify a proposition to the extent that it has one absolute truth-value in all contexts—for if the definition that best serves our purposes is impossible to fulfill, then it is no longer useful and should be thrown out.

### **III. An Answer to the Problem of Faultless Disagreement**

---

With this view of propositions in mind, let us look back at John and Mary's dispute over rhubarb. What proposition is John expressing when he says "rhubarb is delicious"? I believe that the sentence is ambiguous between several different possible propositions, of which these are three:

**P1:** Rhubarb is delicious by John's standard of taste (is delicious from John's context of assessment).

**P2:** Rhubarb has the objective property of deliciousness (is delicious from all possible contexts of assessment).

**P3:** Rhubarb is delicious by some particular standard of taste *A* (for instance, is delicious from Mary's context of assessment).

All three of these propositions have different truth-conditions. P1 is true iff rhubarb is delicious to John. P2 is true iff rhubarb is delicious to everyone (or is delicious independent of anyone's opinions). P3 is true iff rhubarb is delicious by the standard of taste *A*, which should be picked out by John's assertion. For instance, if John is asserting that rhubarb is delicious to Mary, then that proposition is true iff rhubarb is delicious to Mary.

The important thing to remember, which distinguishes this view from both contextualism and relativism, is that John could be expressing *any* of these three propositions (or other possible propositions that I have not thought of). All three are equally legitimate interpretations of the sentence "rhubarb is delicious." Contextualism assumes that John is asserting P1, and that Mary is not asserting the negation of P1, but another proposition, namely that rhubarb is not delicious to Mary. Relativism does not hold that John is asserting any of these three propositions, but rather a proposition that, in its truth-conditions, must collapse into one of the three. The relativist view is that the sentence "rhubarb is delicious" expresses the proposition that rhubarb is delicious, which is true iff rhubarb is delicious in context of use  $C_U$  and context of assessment  $C_A$ . A careful look at these truth-conditions shows that they must collapse into the truth-conditions of either P1, P2, or P3. While there are various contenders for what contexts one commits oneself to their assertion being true in (each of which could be translated into a variant of P1, P2, or P3), MacFarlane says that one commits oneself to their assertion being true in the context in which the asserter

---

is evaluating the challenge (the asserter's context at a later time).<sup>18</sup> The truth conditions for the proposition that rhubarb is delicious are thus equivalent to P1 except with a specified time different from the time of utterance—the time that the assertion is being challenged. All this shows is that P1, P2, and P3 each can be further clarified over parameters like time (which is something I will address in section IV). Thus we can say that the relativist says that when John says “rhubarb is delicious,” he asserts P1\*, that rhubarb is delicious by his standard of taste and will be so at any time his assertion is challenged (this is like saying “rhubarb is delicious to me, and always will be,” seeming to imply that this “always will be” is the only difference between the contextualist and relativist views of what John is asserting). This may not be entirely accurate, given that future contexts in which the assertion will be challenged can vary by other parameters than by time—but the same method can apply to clarifying over those parameters as well. The relativist, like the contextualist, holds that there is one determinate proposition that John is committing himself to the truth of when he says “rhubarb is delicious.” I hold that P1, P1\*, P2, P3, and other interpretations are all equally possible truth-commitments of John's utterance.

Think of it in terms of one of MacFarlane's norms of assertion: “in asserting that *p* at *C*<sub>1</sub>, one commits oneself to withdrawing the assertion (in any future context *C*<sub>2</sub>) if *p* is shown to be untrue relative to context of use *C*<sub>1</sub> and context of assessment *C*<sub>2</sub>.”<sup>19</sup> Thus if Mary interprets John as asserting P2, then she can expect him to withdraw his assertion when she gives him defeating evidence, namely that rhubarb is not delicious to her. Because even from John's context of assessment, P2 is false if there is any context of assessment from which rhubarb is not delicious. But John can very well respond that Mary has misinterpreted him—that what he meant when he said “rhubarb is delicious” was P1 (or P1\*), and so her evidence is irrelevant to the truth of his assertion. This is a possible situation no matter how Mary interprets “rhubarb is delicious”—John can always reply, “that's not what I meant!” rather than be forced to withdraw his assertion. Thus in order for MacFarlane's norm of assertion to apply, we cannot simply assume that “rhubarb is delicious” always has one particular set of truth-conditions. Instead, Mary must ask John, “what do you mean?” before she can evaluate his statement and hold him to the commitment to truth he has engaged in by asserting such.

---

<sup>18</sup> MacFarlane 2005, p.336.

<sup>19</sup> MacFarlane 2005, p.336.

---

What answer does this provide to the problem of faultless disagreement? Two things: first, we cannot know simply from John's statement "rhubarb is delicious" and Mary's "rhubarb is not delicious" whether or not John and Mary are having a genuine disagreement. Both must clarify which proposition they are asserting before we will know whether they are disagreeing: if John asserts P1 and Mary asserts Not-P2, or if John asserts P3 and Mary asserts Not-P1, we can see that they are not disagreeing in any meaningful way. But if John asserts the same proposition that Mary rejects, then we can see that they are disagreeing (sticking to the accept/reject intuition about disagreement that MacFarlane rejects). If context is not sufficient to clarify their assertions and they themselves do not go through an explicit process of clarification, then it is indeterminate whether or not they disagree.

Second, faultlessness and contradiction are genuinely incompatible; there is no such thing as a faultless disagreement. If you accept the laid-out definition of propositions and also accept classical logic, then the truth of P is genuinely incompatible with the truth of Not-P in *any* context. Relativist explanations of faultless disagreement do not work because, viewing propositions as ideally clarified, any explanation where both John and Mary's assertions are true means that John is not asserting the same proposition that Mary is rejecting.

This immediately raises three questions: first, how is this position different from contextualism? Second, what determines which proposition John and Mary express, when they make their respective assertions—is a proposition determined by what they say, or what they mean? And third, how can this position explain our intuition that there are faultless disagreements, if there are in fact none?

This view is different from contextualism because it does not assume that John means P1 when he says "rhubarb is delicious." From a contextualist position, John *must* mean "rhubarb is delicious to me," either through a hidden indexical or the meaning of "delicious." Thus the contextualist immediately concludes that when John says "rhubarb is delicious" and Mary says "rhubarb is not delicious," that John and Mary *cannot* be disagreeing. I say that no conclusion is yet possible: John and Mary may be disagreeing (for instance, if John means to assert P2 and Mary means to reject P2), and they may be faultless (if John means to assert P1 and Mary means to assert that rhubarb is not delicious to her, which is not the negation of P1), but they cannot be both.

Another way to think of the difference between contextualism and this view is that contextualism says of a sentence S that it expresses proposition P (the function from S to P varying with context of use), while this view says of a

---

sentence *S* that it *may* express *either* proposition *P* or proposition *P\**. In fact, John may not know which of the three propositions *P*<sub>1</sub>, *P*<sub>2</sub>, and *P*<sub>3</sub> he is asserting—he may not have thought to distinguish between the three. While, on a contextualist view, to say of John that he does not know what he is asserting seems strange, given that he is asserting one and only one proposition.

This leads to our second question: what exactly is going on when John and Mary make their utterances? How is it determined which proposition they express? The answer has two parts: first, it is not necessarily determined which proposition they each express; second, if it is determined which proposition they each express, that process happens both through context and through explicit clarification. Thus this view can be seen as a type of contextualism, if you take contextualism generally as the idea that context can help to determine meaning (in which case anyone who believes in indexicals is a contextualist of types)—however, it is different from contextualism in several ways already laid out. Context is often sufficient to determine what proposition an assertoric utterance expresses, but if it is not, then we have an ambiguous assertion which needs to go through a process of clarification to be truth-evaluated. Thus, if Mary and John have both just agreed that there is a monadic property of deliciousness that rhubarb either does or does not possess absolutely, then that is sufficient to contextually determine that they are accepting and rejecting *P*<sub>2</sub>. But if they haven't discussed the issue before and are normal people at a dinner table, then context will most likely not be sufficient to determine what proposition is being accepted or rejected.

In this paper I have discussed at length what John and Mary *say* and what they *mean*, seeming to imply that what they mean is the proposition they express (hence why the clarifying question is “what did you mean?” rather than “what did you say?”). This may seem troubling, especially thinking about cases in which what is said and what is meant are radically different, like the example of a man saying “Mary’s husband is kind to her” of a man who is not Mary’s husband, but actually her lover (her husband is actually cruel to her). What the man meant to say is true, but what he said is false. Keith Donnellan draws a distinction between speaker’s reference and semantic reference here, saying that the man said something true of the speaker’s referent (the lover), but something false of the semantic referent (the husband).<sup>20</sup> Examples like this call on me to clarify what I mean when I make the distinction (much as Mary calls upon John to clarify). What I mean by “what John says” is simply the utterance that John

---

<sup>20</sup> Donnellan 1966.

---

makes, not its content. What I mean by “what John means” is what truth-conditional commitments John makes with his assertion. Thus, under this terminology, I take this whole debate to be about what John and Mary mean, not about what they say.

So what is the relation between what John means and what proposition he expresses? If one single determinate proposition is expressed, then what John means and what proposition he expresses will be the same. Thus in asking what John means, Mary is determining which proposition he is asserting with his sentence. What John means, however, may well be indeterminate, as in cases where we say that John does not know which proposition he is asserting. This can happen whenever John has not thought through the truth-conditions he is committing himself to by a particular assertion, as in when someone says that God exists but really has not thought through what exactly God is. There can also be situations in which John has a determinate proposition in mind but fails to express it determinately—perhaps he relies on context when context is not sufficient. John thus leaves himself open to misunderstanding, but this is where the explicit process of clarification comes in, making Mary understand which of the possible propositions expressed by his utterance John means to assert.

The only possible situation that poses a problem for this view is a situation in which there is a determinate single proposition expressed by John’s utterance which is not the same proposition he means to assert. The utterance of “Mary’s husband is kind to her” seems to be a situation like this. It does not seem that one of the possible propositions expressed by this sentence is that her lover is kind to her, even though that is the proposition which is meant to be expressed by the sentence “Mary’s husband is kind to her.” But this is not the case—for “Mary’s husband is kind to her” can mean either that the man who is married to Mary is kind to her (in which case the truth-conditions include that the man referred to is Mary’s husband) or that the man who the speaker supposes to be married to Mary is kind to her (in which case the truth-conditions do not include that the man referred to is Mary’s husband). In the first case, the proposition asserted is false; in the second case, the proposition asserted is true, however, the words used to express it are very misleading. But context prevails—because the man uttering “Mary’s husband is kind to her” is looking at Mary’s lover, standing next to her, it is clear to the people around him that the proposition he means to express is that the man standing next to Mary is kind to her, and those are the truth-conditional commitments he will hold himself to. Thus context provides a possible interpretation of an utterance that may not have been there otherwise. This view of the “Mary’s husband is kind to her” example

---

takes a stance on some controversial issues about definite descriptions, but I think it is a very plausible view, especially taken in terms of the laid-out definition of propositions. The general idea of the relation between utterances and propositions here is that utterances do not express propositions on their own, rather they are clues towards the proposition which is meant by a particular utterance. Thus in a situation where a child pretends it is “opposite day” an utterance of “snow is black” can express the proposition that snow is white. Similarly with the utterance of “Mary’s husband is kind to her.” Thus similarly with the utterance of “a certain guy in the office is a pain in the neck.” This can be a not-entirely-explicit clue towards one proposition meant—that Richard is a pain in the neck; or it can be a more intuitive clue towards another proposition meant—that there is some indeterminate guy in the office who is a pain in the neck. The process of following the clues of the utterance to the proposition meant is the process of clarification, which uses both context and explicit questions like “what do you mean?”

Hopefully this goes a way towards clarifying what exactly is going on when John and Mary make their respective assertions. The distinction between what is said and what is meant and how they relate to propositions, as I take it, is not absolutely crucial to the definition laid out in this paper and the answer it provides to the faultless disagreement problem, though I think it provides the clearest image of what is going on.

Still, the question remains: does this view do better than contextualism in explaining our intuition that there are faultless disagreements? Yes, though it does not give us the satisfaction that relativism gives us by saying that our intuition is correct. Rather, it explains why we make the mistake of thinking that there are faultless disagreements. As we have seen, we are rarely perfectly clear on what we are asserting, and this very often happens because we don’t realize that there are multiple possible different interpretations of an assertion. Thus when John asserts “rhubarb is delicious” we think that he is expressing the proposition that rhubarb is delicious, but we change the truth-conditions of that “proposition” depending on what question we are asked.<sup>21</sup> This is a framing-

---

<sup>21</sup> This does not mean that the content of John’s assertion changes depending on what question we are asked, but rather that our *interpretation* of the content changes depending on what question we are asked. John’s assertion may have determinate content that is failing to be conveyed determinately and thus is open to misinterpretation; or John may not know which proposition he is asserting, meaning that his assertion is indeterminate in content, and he may interpret it as having two different contents depending on framing as



---

effect problem. If we are asked whether John and Mary are having a disagreement, we interpret John as asserting P2 and Mary asserting Not-P2, an interpretation that shows John and Mary as clearly disagreeing. If we are asked whether either John or Mary is at fault, however, we then interpret John as asserting P1 and Mary as asserting that rhubarb is not delicious by Mary's standard of taste. Thus both are asserting something true. Because assertions are often ambiguous we think we are interpreting John and Mary in the same way when we answer the two questions. But on closer examination, it becomes clear that we are not.

But why do we fail to see the ambiguity, falling for the framing effect? I think that when we ourselves are locked in the dispute, we fall for an essentialist intuition, thinking that there is one fact of the matter as to whether rhubarb has deliciousness as part of its essence. Here we are locked in a real disagreement (asserting and rejecting P2), deceived by the thought that our tastes somehow represent a fact of the matter. When we are reminded of the truth that there is no essence of deliciousness or non-deliciousness in rhubarb, *then* we call the dispute faultless, conceding that each may have their own equally legitimate opinions about the matter. But, out of charity to our earlier selves, who engaged in the dispute as if it were a disagreement about a fact of the matter, we interpret the dispute as a genuine disagreement. This same principle of charity operates from an outside perspective: since the parties involved see themselves as engaged in a faultless disagreement, we interpret their assertions in order to best fit their self-view, thus falling for a framing effect that allows us to see the dispute both as faultless and as a genuine disagreement.

Thus falls Wright's Ordinary View. Before we mourn, however, we must remember that it is not the first of ordinary views that philosophy has shown to be internally inconsistent. And in order to rescue the Ordinary View, it is required that we sacrifice our ordinary views about much more fundamental issues: either classical logic (as in Wright's intuitionistic attempt to save the Ordinary View), or truth, disagreement, and propositions (as in MacFarlane's relativistic attempt to save the Ordinary View). And it is much harder to explain why we have been fundamentally mistaken about the nature of truth, logic, propositions, or disagreement for so long than it is to explain our mistake about John and Mary's dispute about rhubarb.

---

well. Hence why John can perceive himself as engaging in a faultless disagreement, also falling for the framing effect.

---

By seeing John and Mary's assertions as ambiguous, we can explain our mistaken intuition of faultless disagreement without being subject to the great philosophical costs implied by truth relativism. We can keep our truth-predicate simple, absolute, and context-independent, keep our strong intuition that two people disagree iff one accepts a proposition that the other rejects, and use a more intuitive and less arbitrary definition of propositions whose truth depends exclusively on properties of and relations between the entities that proposition is about. The only thing that remains to be seen is whether this definition of propositions, which the whole explanation rests on, can be fulfilled.

#### **IV. Is Ideal Clarification of Assertoric Content Possible?**

Another way of putting this question is: is there any context that propositional truth must be relative to? If propositional truth must be relative to some varying context, then the truth-conditions cannot be ideally clarified and a proposition's truth-value cannot be absolute. In addition, MacFarlane's argument will then hold: if propositional truth is already relative to some context, then there is no reason why it cannot be relative to another context (for instance, contexts of assessment). My definition of propositions hinges on the possibility of their absolute truth.

There are a few things that MacFarlane says that propositional truth is already relative to: times, epistemic standards, and possible worlds (the least controversial of the three).<sup>22</sup> I hope to show that it is not only possible but *necessary* to clarify an assertion so that its content does not vary over these parameters. Hopefully these examples will show that the process of clarification can be generalized across parameters, including those which I did not foresee in this paper.

With regards to time, MacFarlane uses the example of the proposition that Joe is sitting. This is distinguished from the proposition that Joe is sitting now, or at any other particular time—rather, the proposition is temporally neutral. MacFarlane says, “If you asserted this proposition at 2 PM and I denied it at 3 PM, we have not in any real sense disagreed. Your assertion concerned Joe's position at 2 PM, while my denial concerned his position at 3 PM.”<sup>23</sup> MacFarlane takes this to be a refutation of the idea that disagreement is composed of one person accepting and the other rejecting the same proposition. But it seems that an assertion at 2 PM that Joe is sitting actually expresses the

---

<sup>22</sup> MacFarlane 2005, p.323.

<sup>23</sup> MacFarlane 2007, p.22.

---

proposition that Joe is sitting at 2 PM, which is not the same as the temporally neutral proposition that Joe is sitting (surely the asserter must believe that Joe is sitting at 2 PM in order to make the assertion at 2 PM, while she need not believe that Joe was sitting at 1 PM or will be sitting at 3 PM). Similarly for an assertion at 3 PM that Joe is sitting—at least by our definition of propositions, the two are certainly not expressing the same proposition. The closest proposition, by our definition, to a temporally neutral proposition is the proposition that Joe is sitting at any possible time that the proposition is evaluated, or in other words, that Joe is always sitting. The truth-value of this proposition, just as the truth-values of the proposition that Joe is sitting at 2 PM or that Joe is sitting at 3 PM, does not vary with time. This means that this proposition is in fact not the temporally neutral proposition which MacFarlane believes can be asserted, the truth-value of which varies with time. Our position is that this proposition does not exist, because anyone who commits themselves to it being true must commit to it being true at any particular time or range of times. If, at 2 PM, you assert the temporally neutral proposition that Joe is sitting, your assertion will be shown to be false at 3 PM, when he is not sitting. The only way your assertion could be false at that time would be if you committed yourself to the proposition that Joe is sitting being true at any time your assertion might be challenged—meaning that you asserted the proposition that Joe is always sitting. If you protest that you did *not* assert that proposition, that the fact that Joe is sitting at 3 PM does not prove your assertion at 2 PM false, then what you are in fact saying is that you asserted the proposition that Joe is sitting at 2 PM. Thus a commitment to the truth of a temporally neutral proposition collapses into a commitment to the truth of an eternal proposition. For if there is any time at which Joe is not sitting, then the temporally neutral proposition can be shown to be false. If Joe is sitting at 2 PM, however, then the proposition that Joe is sitting at 2 PM is still true at 3 PM, when he is not sitting; conversely, the proposition that Joe is not sitting at 3 PM is equally true at 2 PM, when he is sitting. Thus we find not only that it is possible to disambiguate “Joe is sitting” into propositions that have truth-value independently of time, but also that it is necessary to do so in order to evaluate the truth of any particular assertion that Joe is sitting.

As for epistemic standards, the case is exactly parallel to that of standards of taste. Substitute John and Mary for G.E. Moore and a Skeptic. We hear the following dispute between them: Moore says “I know that I have hands,” and the Skeptic responds, “you don’t know that you have hands.”<sup>24</sup> Could this be

---

<sup>24</sup> Richard 2004, p.215-216.

---

another faultless disagreement? Could they be disagreeing over the same proposition that Moore knows that he has hands, which varies in truth-value according to the epistemic standards of the speaker, or of the assessor? By our definition of propositions, no. Moore is either asserting that he knows that he has hands by his own epistemic standard, or that he knows that he has hands by a particular other epistemic standard, or that he knows that he has hands by all possible epistemic standards (parallel to the possible assertions that John could be making by “rhubarb is delicious”). It is necessary that we know which of these he is committing himself to in order to know which proposition he is asserting, and whether or not it is true—similarly with the Skeptic, in order to determine whether or not the two are having a genuine disagreement. Just as with disputes over matters of taste, it is both possible and necessary to clarify assertions into propositions that have truth-value independently of the epistemic standard of the assertor.

MacFarlane’s rejection of the accept/reject definition of disagreement comes back with the conception of propositional truth as relative to possible worlds. This is the most intuitive parameter to which propositional truth could be relative to—it is intuitive and perhaps common practice to say that a proposition is true in some worlds and false in others, or (substituting counterfactual talk for talk about worlds) that a proposition *would have* been true (or false) if things had been different. Here is MacFarlane’s example:

Consider Jane (who inhabits this world, the actual world) and June, her counterpart in another possible world. Jane asserts that Mars has two moons, and June denies this very proposition. Do they disagree? Not in any real way. Jane’s assertion concerns our world, while June’s concerns hers. If June lives in a world where Mars has three moons, her denial may be just as correct as Jane’s assertion.<sup>25</sup>

The question is whether Jane is really asserting the same proposition that June is denying. By our definition of propositions as having one and only one truth-value, it seems that they are not. This leads us to a more specific question: is it possible to take the contextual information of what world a proposition is uttered in and make it an explicit part of the propositional content?

---

<sup>25</sup> MacFarlane 2007, p.23.

---

Yes. Jane can assert truly that Mars does not have (only) two moons in June's world. Or, once again replacing talk of worlds with counterfactual talk, Jane can assert the proposition that if Mars had three moons, Mars would not have (only) two moons. How can we make true statements about possible worlds, if it is not possible to specify a world in the content of the proposition itself?

Indeed, we can see that the sentence "Mars has two moons" is ambiguous in the same way that "rhubarb is delicious" is, merely replacing standards of taste with worlds. "Mars has two moons" can express three possible propositions (with variants):

**J1:** Mars has two moons in Jane's world (the actual world).

**J2:** Mars has two moons in all possible worlds.

**J3:** Mars has two moons in a particular possible world, or set of possible worlds (for instance, June's world).

Thus it seems that it is not only possible, but also necessary to clarify an assertoric sentence into a proposition that has its truth-value independently of the parameter of possible worlds, in order that its truth-value may be assessed.

This may have a somewhat disturbing consequence: that all true propositions are true in all possible worlds. This would lead to the seemingly absurd conclusion that all truths are necessary truths. This conclusion seems absurd because it obliterates the distinction between necessary and contingent truths, seeming to say that it is equally necessary that Mars has two moons and that nine is equal to nine. While our definition of propositions does lead to the conclusion that all true propositions are true in all possible worlds (indeed, in all possible contexts of assessment), this does not require us to eliminate the distinction between necessary and contingent truths, but merely to reformulate it. One way to do this might be to distinguish between the core content and the clarifying content of a proposition. With respect to any varying context or parameter, we can define the core content as that part of the proposition that does not change as we vary across that parameter. Thus the core content with respect to time of the proposition that Joe is sitting at time(s) X is simply that Joe is sitting; the core content with respect to worlds of the proposition that Mars has two moons at world(s) Y is simply that Mars has two moons. It is by taking these core contents to be "propositions" that MacFarlane is able to make

---

“temporally neutral propositions,” or “world-neutral propositions,” the truth-value of which varies depending on the context of assessment. The core content tells you what to look for once you have fixed the value of the varying parameter, picking out truth-conditions within any particular context (in your context, look to see if Joe is sitting, or if Mars has two moons), while the clarifying content picks out the context or contexts in which those truth-conditions should be evaluated in order to assess the truth of the proposition. Contrary to MacFarlane’s assumption, the core content alone (with respect to any particular parameter) is usually not sufficient to truth-evaluate a proposition, because we are capable of asserting things about contexts which we do not immediately occupy (as in when Jane asserts truly that Mars does not have (only) two moons in June’s world). This leads us to the distinction between necessary and contingent truths—necessary truths are those in which the core content with respect to possible worlds (the modal core content, we can call it) *is* sufficient to pick out a single truth-evaluable content. For necessary truths, no clarification over possible worlds is needed—there need be no modal clarifying content. One may add such content, asserting that nine is equal to nine in the actual world, or that nine is equal to nine in June’s world, but while these may change the sense of the assertion, it will not change the truth-value of the proposition uttered. For contingent truths, however, modal clarifying content is needed in order to fix a truth-value for the proposition. We need to know whether “Mars has two moons” expressed J1, J2, or J3 in order to evaluate its truth. Thus the truth of the modal core content that Mars has two moons is contingent on what clarifying content is added; while the truth of the modal core content that nine is equal to nine is necessary. This allows us to talk about necessary and contingent truths while acknowledging that all propositions have their truth-value absolutely and independently of any parameter, including that of possible worlds. Though this definition may thus complicate our talk of modality, it so greatly simplifies our notions of truth and disagreement that I think the balance sheet of philosophical costs still weighs in its favor.

I believe I have shown, then, that a definition of propositions as ideally clarified assertoric content is, in fact, a plausible definition; and that in order to evaluate the truth of certain assertions, it is not only possible, but necessary that the truth of the propositions they express is not relative to the parameters of time, epistemic standards, or possible worlds. I also believe that this method can be generalized to show the same for other unforeseen parameters.

---

## V. Summary and Conclusion

The position I have attempted to argue for in this paper runs thus:

Propositions are a philosophical tool used to rectify the inadequacy of assertoric sentences. This inadequacy is twofold: first, sentences are linguistic form, rather than content; second, sentences are ambiguous. The definition of propositions that best serves the purpose of rectifying this inadequacy (and thus best serves the purposes of philosophy) is the definition of propositions as ideally clarified assertoric content, having one absolute truth-value which does not vary with any parameter, including worlds, times, places, epistemic standards, standards of taste, etc. Given this definition of propositions and classical logic, the Ordinary View of faultless disagreement as put forward by Wright is false—there is no such thing as faultless disagreement. Since ordinary discourse is often unclear and overdependent on context, utterances are often ambiguous in what proposition they express, and this can account for our intuition that there is such a thing as faultless disagreement.

This position itself finds its primary support from our intuitions. Throughout this paper I have appealed to our intuitions about truth, propositions, assertoric commitment, disagreement, and the purposes of philosophy in order to outweigh our single intuition about faultless disagreement. I think this is worth doing, as far as our balance sheet of which intuitions we should sacrifice and which we should keep is concerned. But does that leave my argument without any deeper theoretical support than how we *prima facie* feel about truth, disagreement, propositions, and the lot?

First, I am not sure that anything more than intuitions are needed for this type of argument. The question here is not about the empirically investigated structure of the world, rather it is about our assertoric and philosophical practice—the primary data of evidence in this realm are our intuitions and our practice. And if you look at the arguments for contextualism or relativism, these are based upon intuitions as well.

Second, perhaps I can further assuage doubts by making more explicit an argument that has been used throughout the paper. Our assertoric practice is such that if we are aware of a parameter which might vary the truth-value of our assertion, we implicitly fix a value or set of values to that parameter in the proposition we assert. Consider a caveman and a modern man who both assert “it is noon.” The caveman believes the world is flat, and thus does not understand that time can be relative to one’s position on the earth. The modern man, however, is quite aware of the roundness of the earth, time zones and how they vary with your longitudinal position on the planet. If you say to the

---

caveman that it is not noon many miles away, he will argue with you emphatically, taking you to have disagreed with him. If you say to the modern man that it is not noon in Japan, while his assertion was made in New York, he will agree, not taking you to have contradicted his assertion, for the value of his time zone was implicitly fixed in his assertion. Thus to refuse to set a value for a parameter which you know perfectly well varies the truth-value of your assertion is to violate assertoric practice—to act like the caveman when you have the modern man’s knowledge. Regardless of your intuitions about abstract things like propositions or truth, if you agree with this characterization of assertoric practice, then much of my view will follow. Attempt to imagine a world in which propositions can be neutral or relativized with respect to parameters we very well know vary the truth of those propositions, and I think the assertoric practice in that world would be very different than it is in ours.

There is a final general objection to the definition of propositions as ideally clarified assertoric content: what type of useful tool is it that requires us to clarify over all sorts of parameters that really don’t matter to what we’re talking about? When one asserts that rhubarb is delicious, they’re certainly not thinking about possible worlds, times, epistemic standards, and any other crazy philosophical parameter, though they may be thinking about standards of taste, if they think about it closely. Does this mean that any assertion is ambiguous or false that does not go through the cumbersome process of adding on qualifications like “in all possible worlds and over all times and under this epistemic standard and by my standard of taste”? And indeed, how do you account for everyday discourse, where people almost never explicitly specify over these parameters and yet understand each others’ assertions and evaluate their truth-values with full competence?

My answer to this is that the tool of propositions is not nearly as cumbersome as it seems because all of the process of clarification that we went through explicitly here usually happens effortlessly and under the surface. The work done by context in the translation from sentences into propositions is not to be underestimated. Thus in everyday conversation two people may not even realize that a parameter could vary, they are so accustomed to its assumed constant value—for instance, in the case of everyday common sense, there is a fairly set epistemic standard for “know” which people only fully realize can vary when they begin to dabble in philosophy. It is only in cases like the rhubarb dispute where context is not enough, and explicit clarifications like “by all standards of taste” or “by my standard of taste” are needed in order to truth-evaluate assertions and understand clearly what arguments are about.



---

And it seems that, in cases like John and Mary's dispute over rhubarb, clarification is sorely needed. Without clarification, arguments can take the form of pointless disputes, where both parties are talking past the other and getting absolutely nowhere. This is indeed the final image that MacFarlane paints of the "faultless disagreement," where the challenger and the original asserter, each believing themselves to be utterly correct, are completely unable to get through to the other: "we have all the normative trappings of real disagreement, but without the possibility of resolution except by a relevant change in one or both parties' contexts of assessment." MacFarlane himself admits that "this can look like a pretty silly game."<sup>26</sup> Without clarification, it seems that John and Mary are doomed to have their dinner ruined by a drawn-out, intractable dispute based on a misunderstanding of each others' assertions. Armed with a definition of propositions as ideally clarified assertoric content, however, John and Mary can either see how silly their game is, understanding that they are not disagreeing at all, or, if they genuinely disagree, they can clarify exactly what proposition is in dispute, and exactly what the truth-conditions of that proposition are. Similarly, if philosophers don't want to get locked in silly games based on misunderstandings, and instead engage in genuine and clear disagreements, perhaps the best first step would be to make sure that the propositions they are arguing about have one truth-value, one set of truth-conditions that can be laid out and examined precisely. For the purposes of disagreeing not in a murky, "faultless" manner like John and Mary, but clearly, like philosophers, I believe that the definition of propositions put forward in this paper can be a very useful tool.

---

<sup>26</sup> MacFarlane 2007, p.29.

---

## Works Cited

- Donnellan, Keith. Reference and definite descriptions. *Philosophy Review*, 75: 281-304, 1966.
- Kölbel, Max. Faultless Disagreement, *Proceedings of the Aristotelian Society*, 104: 53–73, 2004.
- MacFarlane, John. Making sense of relative truth, *Proceedings of the Aristotelian Society*, 105: 321–339, 2005.
- Relativism and disagreement. *Philosophical Studies*, 132:17–31, 2007.
- Richard, Mark. Contextualism and Relativism, *Philosophical Studies*, 119: 215–42, 2004.
- Wright, Crispin. Realism, Relativism and Rhubarb, in Greenough and Lynch (eds), *Truth and Realism*, Oxford: Oxford University Press: 38-59, 2006.

---

# Dispositional Properties and Humean Supervenience<sup>1</sup>

JEREMY GOODMAN  
*Brown University*

Abstract: David Lewis' thesis of Humean Supervenience combines the claims 1) that there are no necessary connections between distinct existences and 2) that truth supervenes on being. Contra Lewis, we should adopt a dispositional rather than a categorical theory of property individuation. Moreover, contra the conventional wisdom, such a theory is consistent with claim 1). However, it cannot be made consistent with claim 2) without abandoning the standard semantics for counterfactuals.

## Section 1: Introduction

Late in his prolific career, David Lewis wrote that his "work could be seen in hindsight as a campaign on behalf of 'Humean Supervenience': the thesis that the whole truth about a world like ours supervenes on the spatiotemporal distribution of local qualities."<sup>2</sup> (Lewis 1994:473) Brian Weatherson characterizes Humean Supervenience as the conjunction of two metaphysical constraints: the Armstrong Constraint, which says that truth supervenes on

---

<sup>1</sup> This paper was written a couple of years ago, before I had read Lewis' "Ramseyan Humility" (2009) or Alexander Bird's work on dispositional essentialism. As a result, my current thinking on this topic differs appreciably from the views expressed herein. While for the most part I still endorse the arguments, I now think there are much graver circularity problems for dispositional theories of properties than those I raise here. I've also become skeptical that there even are fundamental properties in any substantive sense (see the penultimate footnote). Thanks to Justin Broackes for many helpful conversations on these topics.

<sup>2</sup> My understanding of Lewis' metaphysical oeuvre is based on his *Counterfactuals* (1973), "New Work for a Theory of Universals" (1983), *On the Plurality of Worlds* (1986), "Humean Supervenience Debugged" (1994), Barry Loewer's "Humean Supervenience" (1996), and Brian Weatherson's *Lewis Blog*. My conception of causal powers has been greatly influenced by conversations with Justin Broackes.

---

being,<sup>3</sup> and the Hume Constraint, which says that there are no necessary connections between distinct existences. Given Lewis' ontology, the Hume Constraint amounts to the claim that the way things are in one area of spacetime has no logical implications for how things are anywhere else.

Much of the debate about Humean Supervenience has been over whether the Hume Constraint is compatible with an adequate analysis of nomic notions, such as the notions of laws of nature, of causation, and of chance. However, the partisans in this debate (i.e. Lewis, et al., and David Armstrong, et al.) contend that properties are individuated merely numerically.<sup>4</sup> They do so because it is generally thought that the leading alternative account of properties, according to which they are individuated by the dispositions (or causal powers) that they confer to their bearers, contradicts the Hume Constraint on its face. I shall call this theory of properties the Dispositional Theory and the Lewis/Armstrong theory of properties the Categorical Theory, although this dichotomy is neither exhaustive nor exclusive. In Section 2 I argue that we ought to reject the Categorical Theory. In Section 3 I argue that a certain form of the Dispositional Theory is in fact consistent with the Hume Constraint. However, I shall argue in Section 4 that this form of the Dispositional Theory is inconsistent with the Armstrong Constraint, so it cannot vindicate Humean Supervenience.

## **Section 2: Against the Categorical Theory**

Consider the properties of fundamental physics; having mass is a good example. According to the Categorical Theory, objects that have mass behave as they do because of the laws of nature which mass figures into. That they should behave this way is a condition of adequacy for any theory of laws—that they explain the characteristic behavior of objects instantiating the properties to which they make reference—, and it is a condition that both Lewisian regularity theorists and Armstrongian governing-law theorists claim to satisfy. Furthermore, both camps take the laws of nature to be contingent: For example, they think that the gravitational constant could have had a different value than it actually does. In other words, they believe that there is a possible world in which

---

<sup>3</sup> The content of this slogan is difficult to unpack, but as a rough approximation I will take it to be the claim that all truths in a world are made true by the way that world is intrinsically.

<sup>4</sup> My discussion should be understood as restricted to those properties that figure in fundamental causal laws, corresponding to Lewis' "perfectly natural" properties. See Lewis (1983).

---

massive objects attract each other with a weaker or stronger force than they do in the actual world. Phrased this way, perhaps intuition is on their side.

But now consider a radically different set of laws. Imagine a world where the gravitational constant is zero, where the law of inertia fails to hold, and where objects spontaneously speed up, slow down, or change direction. Does it make sense to say that the objects in this world *have mass*? I do not see how. Or consider the following world, superficially identical to ours: For every object in our world there is a corresponding object that follows the same spatiotemporal trajectory. However, in this world the laws that govern mass and those that govern charge are reversed with respect to ours.<sup>5</sup> Further, the magnitude of the charge of every object in one respective world and the magnitude of the mass of its counterpart in the respective other world are exchanged in a way that the resulting forces remain of the same strength. Is such a world really possible? Or, to put it another way, is it really distinct from the actual world? I cannot see how it could be.<sup>6</sup>

Neither should Lewis. In *On the Plurality of Worlds*, Lewis argues against the tenability of haecceitism, the most common form of which claims that objects are individuated by their primitive haecceity, or “thisness,” such that there could be qualitatively identical worlds distinguished by their containing different objects, by virtue of their instantiating different numerically distinct haecceities. Such purported worlds are completely analogous to a purported world that is identical to ours in terms of the spatiotemporal distribution of objects, but different from ours by its having different laws, by virtue of its instantiating different numerically distinct categorical properties. Lewis is right that we should reject the possibility of the former worlds, and for the same reasons we should reject the latter.<sup>7</sup> Since it is arguably incoherent to suppose

---

<sup>5</sup> That is, if mass is replaced with charge in Newton’s laws of motion and of gravitation, and charge is replaced with mass in Maxwell’s laws of electromagnetism, then the resulting laws govern the behavior of masses and charges in this purported world.

<sup>6</sup> Lewis (2009) doesn’t find this to be a very big bullet to bite. The bullet strikes me as intolerably large.

<sup>7</sup> Lewis (2009) recognizes this analogy between haecceitism about objects and about properties (which he calls quidditism). However, he argues that haecceitism is in worse shape than quidditism, since it entails that an object can both have and not have a property—albeit in different possible worlds—which doesn’t make sense in Lewis’ metaphysics. However, since, for Lewis, objects have all of their properties essentially, there is no analogous problem for quidditism. Be this as it may, quidditism still seems to be in rather bad shape—see the previous footnote.

---

that any fundamental property's causal powers could have been different from those it actually has, we should take properties to be individuated by these powers.<sup>8</sup>

Does this make the laws of nature necessary, or make properties individuated in terms of the laws of nature which they figure into? People sometimes talk as if this is the case, but I think this is the wrong way of putting things. The above thought experiments show that an object's having a property is nothing more than its having certain causal powers associated with the property. In other words, dispositions are ontologically prior to both laws and properties. Talk of properties is just an imprecise way of talking about the clusters of dispositions that individuate them. Thus, if the property of having mass is individuated by the set of dispositions associated with Newton's laws of inertia and of gravitation, then it is necessary, indeed analytic<sup>9</sup> and trivial, that masses conform to those laws. However, if the property of having mass were instead individuated merely by the disposition to obey the law of inertia, then it would be contingent that masses gravitationally attract one another.

We can of course adjust our terminology to fit our pre-theoretical intuitions, should we have any, about the modal and epistemic status of natural laws, but that way of approaching things misses the point. It is a condition of adequacy on any theory of laws that propositions about the world's nomic character have cognitively and empirically significant content about the nature of the world and the objects in it; but there is no condition of adequacy on theories of laws as to how they concern the nature of properties. Indeed, once we adopt the Dispositional Theory, we should no longer, on pain of vacuity, conceive of laws as stating facts about the nomic powers of properties. Rather, we should conceive of them as generalizations about the properties that happen to be instantiated in the world. If it is indeed a conceptual truth that masses behave gravitationally, then the law of gravitation should be understood as the claim that the world is full of masses. This claim accords nicely with our scientific practice,

---

<sup>8</sup> I believe I have shown that it is incoherent to suppose that a property could have causal powers *wildly* different from those it actually has. However, might it be possible for a property to have *mildly* different powers? One might argue by analogy that, although I couldn't have been a fried egg, I could have been an inch taller than I actually am. This thought leads naturally to the idea of extending the Kripke-Putnam essentialist semantics for natural kind terms to theoretical terms. I strongly prefer Lewis' account, according to which theoretical terms are implicitly defined by their roles in the theories in which they appear. See his "How to Define Theoretical Terms" (1970).

<sup>9</sup> See the previous footnote.

---

since physicists take fundamental properties to be defined by the laws which they figure into, and physicists identify objects as having those properties by virtue of their causal interactions. This practice is perhaps the root of our intuitions that tell decisively against the possibility of worlds in which properties figure into laws significantly different from those in which they actually figure; in turn, these intuitions tell decisively against the Categorical Theory.

So far, there is nothing that should upset Lewis. In fact, given that he is such a vocal defender of functional analysis of higher-level properties, of pre-philosophical terminological practices, of anti-haecceitism, and of regularity theories of laws, one might think he would be inclined to side with the Dispositional Theory rather than the Categorical Theory.<sup>10</sup> Moreover, the Dispositional Theory does not threaten Lewis' ability to appeal to the objective naturalness of certain properties.<sup>11</sup> In whatever sense it might be thought legitimate to think of some categorical properties as natural and others as gerrymandered, it seems just as legitimate to think of there being natural and gerrymandered clusters of dispositions.

### Section 3: The Dispositional Theory and the Hume Constraint

Lewis' main issue with the Dispositional Theory is that it seems to flatly contradict the Hume Constraint.<sup>12</sup> If fundamental, intrinsic properties are essentially dispositional, then it seems that there must be necessary connections between the character of distinct spacetime regions, particularly between an object's past and future locations. Perhaps this is true for some notions of dispositional properties, but I do not think it is true for all.

Consider a fragile glass. It is plausible to say, given the Dispositional Theory, that when we ascribe fragility to a glass we are ascribing to it the disposition to break when struck.<sup>13</sup> It is natural to think that putting intrinsic dispositional properties such as fragility in the world violates the Hume Constraint by (1) making it necessary that a fragile glass, once struck, must then break, yet (2) according to the Hume Constraint, a fragile glass could be fragile, be struck, not break, and remain fragile. However, the conflict between these

---

<sup>10</sup> Lewis himself thinks that he could adopt the Dispositional Theory if he wanted to. See Lewis (2009).

<sup>11</sup> Lewis thinks such appeals can solve a myriad of philosophical problems; see Lewis (1983).

<sup>12</sup> This is the opinion of all of the authors cited in footnote 2.

<sup>13</sup> Actually, we are ascribing to it the disposition, *ceteris paribus*, to break when struck. For simplicity, I will hereafter pretend that there is no such *ceteris paribus* clause.

---

two claims can be resolved if we assume that dispositions are triggered instantaneously.<sup>14</sup> We must read the “then” in (1) as expressing a conditional rather than a temporal relation, and “remain” in (2) as “fragile both before and after the glass was struck.” It is easy to see that the following claims are compatible: (1\*) if at time  $t$  a fragile object  $O$  is struck then it must break at time  $t$  and (2\*) it is possible for  $O$  to be fragile immediately before and after  $t$ , be struck at  $t$ , and not break at  $t$  (by, arbitrarily, ceasing to be fragile at the very instant  $t$ ). The Hume Constraint and the Dispositional Theory are compatible after all.

Two further points. First, one might be somewhat suspicious of (2\*). Here the categorical way of speaking might actually be helpful, since, although it ultimately rests on a conceptual confusion, in some cases it more closely mirrors our everyday talk. It should be clear from the previous section that talking of an instance of a categorical property behaving in accordance with a natural law is equivalent to talking of an instance of a dispositional property behaving as it must by definition.<sup>15</sup> Thus, if we think of fragility as a categorical property and of the glass as existing in a world governed by a law to the effect that fragile objects break when struck, it seems to be a perfectly coherent possibility for the fragility to blink off at the moment the glass is struck and blink back on immediately afterwards, without the glass ever breaking, just as a camera flashes just at the instant when a picture is taken, and therefore never records a dark world. If dispositional talk is always equivalent to certain categorical talk, then we can use the idea of blinking categorical properties to get a conceptual handle on the blinking dispositions in (2\*).

Secondly, just as the Dispositional Theory shows that the nomic necessity associated with property-instances obeying natural laws is a species of conceptually grounded metaphysical necessity, so does the account of instantaneously exercised dispositions reveal the nomic necessity associated with dispositions to be conceptual and metaphysical as well. The impossibility of a disposition and its trigger being co-instantiated without the disposition simultaneously being exercised is grounded in the very meaning of the concept of dispositions (so understood). This impossibility is not like the proposition that

---

<sup>14</sup> This seems the natural thing for the Humean (and perhaps the physicist, too) to say regardless.

<sup>15</sup> Justin Broackes offers compelling arguments for this equivalence in his “The Autonomy of Colour” (1992).



---

nothing can travel faster than light; rather, it is like the proposition that nothing can be both round and square.

I do not know whether this account of instantaneously exercised dispositions can be made to work for all dispositional properties that might threaten the Hume Constraint. However, I will hereafter assume that it can, since my goal is to expose a deeper difficulty about the compatibility of the Dispositional Theory and Humean Supervenience that concerns the Armstrong Constraint rather than the Hume Constraint.

#### **Section 4: Nomic necessity**

Up to this point I have not said what Lewis' Humean analysis of laws consists in. It is a regularity theory, borrowed from Frank Ramsey. Imagine we were somehow able to know and synthesize all of the facts about our world, and arrive at a set of postulates that describes the totality of these facts with the optimal balance of simplicity and explanatory strength. Taking these postulates as the axioms of our scientific theory, the natural laws of our world are defined as the theorems that follow deductively from these axioms. This account arguably captures the great majority of our pre-theoretical intuitions about laws.<sup>16</sup>

Laws incontrovertibly entail generalizations. For example, the claim "it is a law that all Fs are Gs" entails the generalization "all Fs are Gs." However, it is also widely agreed that the converse is not true: Laws are more than mere generalizations. Although not all true generalizations are laws on the Ramsey-Lewis account, critics of regularity theories claim that the account nonetheless ascribes lawhood too widely: That it fails to capture an essential ingredient needed to make a generalization a law. To bring the worry into focus, consider two worlds each consisting of a glass in a vacuum, and nothing else. Following Lewis, let us engage for a moment in categorical talk. Suppose that neither world's glass is ever struck or ever breaks. Both glasses are qualitatively (i.e. categorically) indistinguishable. However, in one of the worlds, it is a law that glasses are fragile; in the other world it is not a law that glass is fragile. Intuitively, there are such distinct worlds. What distinguishes them—that is, what makes it the case that in one the glass is fragile—is the truth of the counterfactual that, if the glass were struck, it would break. However, since Lewis is committed to the Armstrong Constraint and thinks that the worlds described must be intrinsic duplicates, he is forced to bite the bullet and claim

---

<sup>16</sup> See section 3.3 of *Counterfactuals*, and Loewer (*op. cit.*) for a compelling defense of this claim.

---

that the two worlds I just described must be one and the same. There is nothing he can appeal to about one world to make the counterfactual true that would not apply just as well to the other.

Before assessing the implications of this result for Lewis' view, there are a number of lessons to draw. The first is that dispositions are grounded in *de re* counterfactuals: If *x* is disposed to do/undergo *y*, given trigger *z*. Just in case were *z* to obtain for *this very x*, *x* would do/undergo *y*. Second, the Armstrong Constraint requires that the intrinsic nature of the world make these counterfactuals true. Third, the truth of these counterfactuals seems to provide the extra "modal oomph" necessary for a generalization to count as a law. For convenience, I will assume that it is a sufficient condition as well. That is, I shall assume that the proposition, "It is a law that all *F*s are *G*s," is true if and only if the proposition, "For every object, were it *F*, it would be *G*," is true.

Let us now return to the case of the two glasses in the void. It does intuitively seem, contra Lewis, to describe two distinct worlds; this counts somewhat strongly against the Ramsey-Lewis regularity theory. But what if Lewis were to adopt, as I advocated in Section 2, a dispositional account of intrinsic properties (which, as I argued in Section 3, is consistent with the Hume Constraint)? Such a dispositionalist regularity theory seems to trivially avoid the objection: Lewis could simply say that one world's glass has the intrinsic dispositional property of fragility, while the other world's glass does not. He can ground our intuitions about laws by appealing to dispositional intrinsic properties. These dispositions are further grounded in certain *de re* counterfactuals. Since Lewis has a reductive theory of the semantics of counterfactuals and of *de re* modality in his metaphysical bag of tricks, he is finally home free.<sup>17</sup> Humean Supervenience emerges victorious.

This Humean reply, however, is too clever by half. The Dispositional Theory, though it can be made compatible with the Hume Constraint, is incompatible with the Armstrong Constraint. The Dispositional Theory says that the intrinsic nature of the objects in the world is dispositional. These dispositions are grounded in certain *de re* counterfactuals. But what grounds these? According to Lewis, a counterfactual is true in a world just in case in the world most similar to it in which the counterfactual's antecedent obtains, its consequent obtains as well. Further, *de re* modal claims about an object in a world are true in another world just in case that world contains a counterpart of the object and the claim is true of this counterpart at that world. But what determines the similarity

---

<sup>17</sup> See *Counterfactuals* and Chapter 4 of *On the Plurality of Worlds*, respectively.

---

relation between possible worlds, and the counterpart relation between possible objects? The answer, according to Lewis, is the similarity between the possible worlds/objects with respect to their intrinsic character.<sup>18</sup> But now we have run in a circle. The Dispositional Theory makes the intrinsic nature of things dispositional, dispositions must be grounded in *de re* counterfactuals, *de re* counterfactuals must be grounded in similarity, and similarity must be grounded in the intrinsic nature of things. This circle certainly seems vicious, since, at least to my mind, it renders its elements unintelligible. The Armstrong Constraint is violated; there is not enough being to ground truth.

To get a better feel for the problem, consider a well-behaved world, such as (presumably) our own, using categorical talk. Unlike the glass-in-the-void worlds, presumably there are more than enough regularities for the Ramsey-Lewis method to generate generalizations that correspond to what we intuitively think are this world's actual laws. The question is whether or not, on Lewis' account, they also have the modal oomph required for genuine lawhood; that is, taking these generated generalizations to be *material* conditionals that apply universally, does the Lewis semantics for counterfactuals validate the corresponding universally applying *subjunctive* conditionals, too?

I think there is good reason to think that it does.<sup>19</sup> Consider some concrete cases. The Ramsey-Lewis method plausibly generates the following generalizations about the actual world: All objects have an inertial mass equal to their gravitational mass; medium-sized objects of sufficient density fall when dropped on Earth; extremely hot objects emit blackbody radiation in the visible spectrum; all black holes are singularities; etc. It seems plausible that the corresponding universally quantified *de re* counterfactuals are also true: for each actual object, were it harder to move/denser and dropped on earth/extremely hot/a black hole/etc., it would be proportionally more gravitationally attractive/fall/glow/be a singularity/etc. For example, the fact that it is a theorem of our best physical theory of the actual world that all extremely hot objects glow does seem to be enough to ensure that in the world most similar to our own in which my keys are extremely hot, they glow. Thus, according to the Lewis semantics, if my keys were extremely hot, they would glow. This seems to be the case for objects in general—not just my keys—and for theorems of physics in

---

<sup>18</sup> The extrinsic character of objects is also obviously relevant.

<sup>19</sup> See Loewer (*op. cit.*). Note that this does not depend on the claim, which Lewis denies, that all counternomological worlds are less similar to a given world than the most dissimilar world that nomologically agrees with it.

---

general—not just those pertaining to thermal radiation. If Lewis can indeed get the relevant *de re* counterfactual generalizations to come out true, then perhaps his regularity theory is at least adequate for *our* world.<sup>20</sup>

It should now be clear why Lewis needs the Categorical Theory to reconcile the Armstrong Constraint and the Hume Constraint. Any adequate account of nomic necessity must include laws that support counterfactuals. For this to be consistent with the Armstrong Constraint, these counterfactuals must be made true by the way the world is intrinsically. The Hume Constraint rules out any appeal to laws as basic intrinsic features of worlds, *à la* Armstrong. But if the counterfactuals can be grounded via a similarity relation between possible worlds, and this relation can be grounded in the Humean mosaic of local categorical properties, then there is at least a possibility that Lewis' project can succeed. But once we are pushed to adopt the Dispositional Theory the whole picture collapses. A mosaic of local dispositional properties requires counterfactuals to support the dispositions. But now, unlike the Categorical Theory story, we can no longer appeal to the mosaic to ground these counterfactuals, on pain of circularity. Appealing to categorical properties allows Lewis (rightly, in my view) to reject haecceities as a metaphysician's superstition, by providing a way for him to instead analyze *de re* counterfactuals in terms of categorical intrinsic similarity. What he failed to appreciate is that categorically individuated properties are effectively haecceitary; they too are a superstition worth rejecting, since they conflict with our intuitions about properties' possible causal powers. But, unfortunately, the combination of the Dispositional Theory of properties, the Armstrong Constraint, and the standard semantics for counterfactuals is too austere a metaphysical framework to be adequate. One of these three theses must go.<sup>21</sup> The upshot of this paper is that

---

<sup>20</sup> This would be enough for Lewis, since he does not think we should put much stock in our intuitions about glass-in-the-void worlds; see Lewis (1994).

<sup>21</sup> My money is on the Dispositional Theory. I believe there is a third, deflationary theory of fundamental properties that allows us to save Humean Supervenience. At the fundamental level, Lewis' theory has two ontological ingredients: spacetime and properties. The dispositional theorist has to have three: spacetime, dispositions, and the things that have the dispositions (e.g. particles). The dispositions cause trouble for the Armstrong Constraint, and I think we can do without them. Roughly, the idea is this: Some spacetime coordinates are populated by particles. There is an electron at a coordinate *if, and only if, it lies on an electron-shaped spacetime path of particle-populated coordinates* (where "electron-shaped" is ultimately to be cashed out by Ramsifying a physical theory). Of course, real physics might require a richer ontology, but our toy-

---

the Hume Constraint is not, as is often supposed, an Achilles heel of Lewis' project.<sup>22</sup>

---

physics ontology can serve as a foundation for Lewis' metaphysical program, which is all Lewis really wants from Humean Supervenience.

<sup>22</sup> Of course, it is not at all obvious why we should care about the Hume's Constraint in the first place. Weatherson suggests what he calls the "Sinatra Argument—if we can make it there, we can make it anywhere!" In other words, if we can find a place for folk notions in the austere Humean mosaic, then no doubt we can assimilate them into whatever supervenience base our ultimate theory of fundamental physics ends up providing. See his "Why Care About Humean Supervenience" in Weatherson (*op.cit.*).

## WORKS CITED

Broackes, Justin. "The Autonomy of Colour." *Reduction, Explanation, and Realism*. Ed. David Charles and Kathleen Lennon. Oxford: Clarendon P, 1992. 421-65.

Lewis, David K. *Counterfactuals*. Grand Rapids: Blackwell Limited, 1973.

-----"How to Define Theoretical Terms." *The Journal of Philosophy* 67 (1970): 427-46.

-----"New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (1983): 343-77.

-----*On the Plurality of Worlds*. Grand Rapids: Blackwell Limited, 1986.

-----"Humean Supervenience Debugged." *Mind* 103 (1994): 473-90.

-----"Ramseyan Humility." *Conceptual Analysis and Philosophical Naturalism*. Ed. David Braddon-Mitchell and Robert Nola. Cambridge: MIT Press, 2009. 203-22.

Loewer, Barry. "Humean Supervenience." *Philosophical Topics* 24 (1996): 101-27.

Weatherson, Brian. Weblog posts. *Lewis Blog*. 2005-2006. 31 Mar. 2009 <<http://lewisblog.weatherson.org/>>.

# Bringing a Text to Life: The Role of the Reader in Plato's *Phaedrus*

MELISSA TAN

*University of Chicago*

Abstract: In the *Phaedrus*, Socrates criticizes writing as non-living and deceptive. He later also claims that a good writer will write only for the sake of self-amusement. These apparent indictments of the written word seem to be at odds with the fact that the *Phaedrus* is itself a written text, to which Plato has presumably devoted some care and effort. I will show, however, that Plato uses these claims ultimately to suggest that the reader is responsible for transforming a written text into a dialogue with the text's author. I argue that Plato gets this message across via deliberate but not unsubtle flaws in Socrates' arguments and by highlighting the frivolity of written words, thereby directing the careful reader to recognize the significance of what Socrates leaves unsaid. For Plato, what is left unsaid is a more reliable vehicle for conveying some understanding of reality and truth than mere written words.

Near the end of the *Phaedrus*, Socrates criticizes writing for three main flaws. First, writing makes its users rely on something external. Second, it can at best convey only an appearance or image of reality. Third, it cannot answer questions, defend itself, or choose its audience. In his discussion of these three main flaws, Socrates implies that the reader is merely a passive recipient of the negative effects of writing. He does not explicitly offer any suggestions for what the reader could do to guard against or negate the effects of reading a written work. Socrates' failure to say much, if anything, about the role and responsibilities of the reader may lead to the appearance that Plato is less concerned with the reader than he is with the writer of a text.

However, I will argue in this essay that, through this conspicuous lack of discussion of the role of the reader, Plato communicates that it is the reader who is responsible for breathing life into an apparently non-living text. Specifically, the reader has the ability and responsibility to convert an apparently one-way speech, e.g. a printed essay, into a two-way dialogue. Moreover, the reader's response to a text is crucial in determining whether the text becomes a dialogue. I

---

will show that even though Socrates is explicitly criticizing the deficiencies and negative effects of writing, the careful reader is meant to perceive that Plato is using conspicuous deficiencies in Socrates' criticisms to point us to his actual concern: the reader's response to the text. I propose that Plato does not explicitly state his intended message because he wants us to realize that we need to spend time and effort reading the *Phaedrus* in order to be rewarded with understanding – this is Plato's way of amusing himself and at the same time enabling his written text to effectively weed out the parts of its audience that Plato would consider unsuitable.

Before I go on to my argument, I will first explain the three main flaws that Socrates attributes to writing, and then demonstrate that Plato is actually concerned with the reader's response to the text and not just with the text's effect on the reader. The first flaw Socrates points out is that a reader will start to rely on something outside himself in order to remember information and will thus become forgetful. To demonstrate this point, Socrates uses the story of the god Theuth presenting writing to Thamus, the king of Egypt, who tells Theuth that writing "will introduce forgetfulness into the soul of those who learn it: they will not practice using their memory because they will put their trust in writing . . . instead of trying to remember from the inside, completely on their own" (Plato, *Phaedrus* 275b). I assume that by "those who learn it", Socrates is referring not only to the writer but to the reader as well, who has to learn to interpret the symbols used by the writer. In other words, users of writing – both writers and readers – will no longer devote time and effort to learning something properly but instead rely completely on written texts to store their knowledge. Although Socrates does not explicitly elaborate, I infer that Socrates would say that the user's "trust in writing" is detrimental because writers and readers will grow lazy and dependent on perishable objects that are "external" to the body and soul (*Phdr.* 275b). They will not work to cultivate knowledge that is "truly written in the soul", i.e. ingrained upon the user's soul and based on a true understanding of reality (*Phdr.* 278a).

Here, I want to clarify that when Socrates uses the phrase "remember from the inside", he is not talking about simply being able to recite information from memory (*Phdr.* 275b). Instead, to be able to "remember from the inside", the learner has to gain an understanding about truth and reality. After gaining such an understanding, the learner would presumably not need to memorize any more data. To illustrate this point: If I were trying to learn chemistry, memorizing the results of all the chemical reactions I could think of would not help me truly understand how chemistry works. I would know that A and B



---

together make C, but I would not know how or why. However, if I truly understood the principles underlying chemistry, it is likely that its fundamental principles would start to seem so natural to me that my knowledge of chemistry would no longer depend on how well I have memorized chemical facts. Without being told beforehand what A plus B would make, for instance, I could infer the result based on the principles that I have internalized.<sup>1</sup>

The second main flaw of writing, according to Socrates, is that it can only be an image of reality. Because writing can convey only a semblance of reality but not reality itself, Socrates says, the words of a written text cannot give readers any understanding of truth or reality. He says that readers mistakenly believe that a text can tell or has told them something true about reality, when in fact it cannot. To illustrate this point with an analogy: We look at pinned butterflies in museums and think that we have somehow become more knowledgeable about different butterflies, but a pinned butterfly does not allow us to find out significant things about the true nature of the butterfly, e.g. how this particular species flies. Socrates goes on to say (in the guise of Thamus) that because writing cannot ensure that its audience is suitable – an issue I will discuss further when I reach Socrates’ third criticism – readers can “hear many things without being properly taught, and they will imagine that they have come to know much while for the most part they will know nothing” (*Phdr.* 275b). They will then “merely appear to be wise instead of really being so” (*Phdr.* 275b). In other words, we can read many texts without properly understanding them, and then believe that we have understood the texts and are therefore wise with respect to them when we actually have not understood them at all. Also, since written texts are just representations of reality, even if we do understand a text, the most we can actually grasp is the image that the text presents to us, which is not reality itself.

Furthermore, Socrates then notes that, since written texts can only convey images of reality, anyone who thinks that knowledge of an art can be clearly conveyed through “written instructions” is wrong (*Phdr.* 275c). “Words

---

<sup>1</sup> To internalize something, I suggest, is not merely to commit it to memory but to know it so well that it becomes almost second nature to view the world using that principle. For example, we learn in school that positive and negative electric charges attract each other. Before we are given this fact, it is not something natural to us; however, this knowledge is now ingrained within us (hopefully) such that if we knew A was positively charged and B was negatively charged, we would automatically think that there was an attractive electromagnetic force between them. Internalization of knowledge bears further discussion in a separate essay.

---

that have been written down”, he says, cannot “do more than remind those who already know what the writing is about” (*Phdr.* 275c). In other words, if I were reading a text that described an art with which I had no experience whatsoever, I would not be able to truly understand the art. One example that illustrates this point is instruction manuals. A written instruction manual for operating a camera, for instance, tells me how to change shutter speeds. However, although the manual teaches me this piece of knowledge that is necessary for me to take photographs, the manual is insufficient because the knowledge that it offers me cannot capture the reality of using a camera. When I am confronted with a situation that the manual fails to describe I will be at a loss for what to do. I would have to acquire more firsthand experience using the camera before I could know how to use the camera in non-textbook situations. Once I have acquired this experience, the instruction manual becomes a memory refresher. Socrates claims that since writing can never teach readers about truth or reality, a writer who “knows what is just, noble, and good” (*Phdr.* 276c) should write things down only “for the sake of amusing himself” and “storing up reminders for himself ‘when he reaches forgetful old age’” (*Phdr.* 276d). I will discuss this in more detail later.

The third main flaw Socrates points out is that a written text is inert and non-living: It cannot answer questions, discriminate between audiences, or defend itself against criticism. He compares a piece of writing to a painting, saying that “the offsprings [sic] of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words” (*Phdr.* 275d). Here, Socrates is saying that although I can ask a living person a question and receive her answer through some medium – voice, text, gesture, etc. – the same does not hold for a piece of writing. I cannot ask a written text a question and get an answer in return. He also notes that “if you question anything that has been said because you want to learn more, it [the writing] continues to signify just that very same thing forever” (*Phdr.* 275d). In other words, Socrates suggests that because a written text is fixed and static, the reader cannot access any more information than what is already conveyed by the unchanging words of the text. The text can only repeat itself over and over again, and cannot offer anything additional.

Moreover, Socrates notes that a written text cannot choose its audience, whereas a living speaker can. A written text “rolls about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn’t know to whom it should speak and to whom it should not”, according to Socrates (*Phdr.* 275d). In other words, once an idea has

---

been written down on a moveable object, it can be transported to various places and be read by many different people who may or may not understand it. While a living speaker can restrict conversations to people he thinks will understand him and employ different arguments to suit specific listeners, the writer of a text is unable to choose his audience. The most the writer can do is try to limit his text's circulation to a select few, but such an endeavor may not succeed. Socrates implicitly suggests that a text should not be read by people who do not understand it, because such people would possibly misread it and end up "unfairly" attacking the text (*Phdr.* 275d). In such a situation, Socrates notes, the text cannot defend itself against its detractors' allegations since it has no audible voice. The text will "always needs its father's support" – e.g. further explanation or refutation by its author – which may not be immediately available if the text has circulated beyond the physical reach of its creator (*Phdr.* 275d). Citing the above reasons, Socrates says that writing is not "legitimate": "Legitimate" speech would be "written . . . in the soul of the listener", be able to "defend itself", and "[know] for whom it should speak and for whom it should remain silent" (*Phdr.* 276a).

At this juncture, I want to clarify the distinction between living speech and non-living speech. I suggest that what makes speech living, for Socrates, is the direction of flow of information: Living speech is effectively a dialogue, or a two-way exchange of content, whereas non-living speech involves only a one-way transmission of material. Although it may appear at first glance that Socrates would categorically consider a spoken speech to be living and a written text to be non-living, based on his second criticism of writing, I propose that this is not the case. For instance, it is likely that Socrates would say that a lecture would be effectively non-living if the audience is forbidden from asking questions and responding to the lecturer. It is clear, then, that Plato thinks that the difference between living speech and non-living speech does not lie in the physical location of the speech, be it inscribed on parchment or within the soul of the audience. Thus, the possibility for written speech to be living is not ruled out. I will go on to claim that Plato actually wants us to realize that the livingness of speech depends on the reader's response to it. Specifically, although written speech may seem to be merely a non-living, one-way transmission, the reader has the power and the responsibility to convert it into a two-way dialogue, and in this way breathe life into the speech.

The first indication that Plato wants us to realize that the reader's response is crucial is that even though Socrates explicitly talks about the negative effects that writing has on the reader, we can see that all his criticisms are also

---

concerned with how the reader responds to the text. To repeat, Socrates' three main criticisms are: first, writing causes forgetfulness in its users; second, writing can only convey an image of reality and not reality itself; and third, writing is inert and effectively non-living. I suggest that Plato uses deficiencies in Socrates' three criticisms to show us that the way the reader responds to the text is important.

Socrates' first criticism fails to consider that writers and readers will grow dependent on a text only if they choose to let themselves do so. I acknowledge that there is a temptation to slothfully rely on texts to store knowledge. I do not have to learn chemistry's fundamental principles if I can get all the information I want from a textbook. However, this temptation can be resisted if the writer or reader responds to the text in a certain way. For instance, the author of a textbook presumably retains all her knowledge of fundamental principles even after she has written the textbook; a reader of that book could, with enough effort and experience, come to an understanding of those principles such that the book is no longer needed. From Socrates' conspicuous failure to mention the active role that the reader can play, I propose that Plato wants us to see that the negative effect of forgetfulness actually arises from the way the reader chooses to respond to the text, and thereby realize that the reader is not merely a passive recipient of the effects that writing produces.

Turning to Socrates' second criticism, I do not disagree that writing can convey only an image of reality to readers. In making this criticism, Socrates is concerned with the negative effect that a written text has on its readers: They will be misled into thinking that they understand something about reality when they in fact do not. However, we see from other Platonic dialogues that even a dialogue with Socrates can fail to produce understanding about reality. For instance, at the end of the *Ion*, Ion is just as clueless as he was at the beginning.<sup>2</sup> I suggest that

---

<sup>2</sup> Throughout his dialogue with Socrates, Ion appears to come to no real understanding of or conclusion about whether his skill as a rhapsode is due to his technical mastery or some divine inspiration. Ion holds in the beginning that he does have some technical knowledge, but easily wavers and continually allows Socrates to poke fun at him right up to the end. He allows Socrates to say that "the god deliberately sang the most beautiful lyric poem through the most worthless poet" (Plato, *Ion* 534e); and that "you [Ion] make many lovely speeches about the poet without knowing anything" (*Ion* 542a). At the end of the dialogue, it seems that Ion happily agrees that he is ignorant only because Socrates frames it as a choice between being ignorant and "divine", versus being technically knowledgeable and non-divine. Ion picks the easier former position, and the reader is

---

Plato intends us to realize from this that the way the reader chooses to respond to a speech is much more important than its physical format, i.e. whether it is spoken or written. Not everyone who converses with Socrates and participates in a two-way dialogue will gain understanding. At the same time, not all readers will read a text and mistakenly believe that they have then understood something true about reality. If I approach a text with the knowledge that words can only capture one slice of reality and not the whole of it, I can avoid falling into the trap of believing that after reading the text I will understand something true about reality and become wise.

The above implies that as a reader, I have the power to choose whether a text affects me in a living or non-living way, which brings me to Socrates' third criticism: that written texts are inert and non-living. Socrates claims that a written text is non-living in the sense that it cannot answer questions, defend itself against criticism, or discriminate between audiences – it is merely a one-way transmission of information and cannot provide a two-way dialogue between author and audience. Socrates says that if anyone asks written words anything, the words "remain most solemnly silent" (275d). However, I suggest Plato intends the reader to see that it is not impossible to ask questions of a text and reasonably predict its answers, as long as these questions are relevant to the text. In this way, the text will not "signify just that very same thing forever", since I can learn more about the subject of the text by careful inference (*Phdr.* 275d). For example, if I were to ask Socrates what he would say about academic lectures, I could predict that he would consider lectures to be non-living speech if the lecturer ignores any audience reaction and does not allow the audience to speak. Caveats apply: If the text were nonsensical or badly written, I would not be able to reasonably infer the answer to any question I posed (Mosher). Additionally, it is not true that a text cannot defend itself against detractors without the help of living supporters such as its author. I suggest that a text can defend itself<sup>3</sup> in the sense that we, as readers, can rationally infer the ways in which the author of an essay should<sup>4</sup> respond to objections brought against it,

---

slightly disappointed that Ion does not make any sort of effective stand or try to directly engage Socrates (*Ion* 542b).

<sup>3</sup> I acknowledge that a text technically cannot audibly defend itself since it is not a biologically living thing with agency, but I will continue to use the verb 'defend' for the sake of convenience.

<sup>4</sup> Footnoted because a discussion of the following is beyond the scope of the surrounding paragraph: What the author should say given what we can rationally infer from the text might differ from what the author would actually say (Lear). For instance, Aristotle says

---

given the ideas that he has already expressed in the text. Again, the same caveats apply. If a text were so badly written that even an astute and careful reader could not make out the structure or content of any sort of argument, it is safe to say that it would be quite difficult to find defenses within the text. Finally, even though a written text can be read by anyone with physical access to it, I argue that the author of a text does not completely cede control over the selection of his audience. The author does have ways of choosing his readers, and I suggest that Plato makes use of some of these in the *Phaedrus*.

I suggest that Plato uses two methods to control his audience, such that the *Phaedrus* as a text will “[know] for whom it should speak and for whom it should remain silent” (*Phdr.* 276a). The first method is to have Socrates fail to explicitly discuss the reader’s role in responding to a text. I have already discussed this above. Socrates’ neglect is not immediately obvious, which means that the reader has to pay close attention to what Socrates says and, more significantly, does not say. This weeds out readers who are not paying close attention.

The second method Plato uses is to claim that a writer who knows what is “just, noble, and good” (*Phdr.* 276c) will only write “for the sake of amusing himself, storing up reminders for himself” (*Phdr.* 276d), and thereby to imply that the *Phaedrus* was written frivolously, i.e. not seriously. Superficially, this seems to the reader as though Plato is effectively inserting an escape clause in the text that will allow him to get away with saying anything he wants. This is because we usually assume that if someone tells us that he is not being serious about X, he is saying that he may lie, joke, embellish, and so on with respect to X, and does not want us to believe him completely. The reader also assumes that Plato would probably consider himself to be a writer who is aiming at what is “just, noble, and good” (*Phdr.* 276c). Thus, hearing Plato say that such a writer

---

that “the function of man is an activity of soul which follows or implies a rational principle” (Aristotle, *Nicomachean Ethics* 1098a6). We can rationally infer from this that Aristotle should say that all human beings are equal. However, Aristotle actually says that some people are naturally slaves (Aristotle, *Politics* 1254a22). I agree that when we, as readers, come up with ways in which authors are likely to respond to objections, we are rationally inferring what the author should say. This suggests that the only dialogue that matters when we question a text and engage with it is what is rationally consistent with the author’s writing. In making rational inferences from a text, the reader is taking responsibility for determining what is rationally consistent and then providing a rationally consistent inference – which implies that the reader has the ability and responsibility to respond to a text rationally.

---

will not “be serious about writing” makes us think that Plato could be, for example, covering up his real opinions (*Phdr.* 276c). The reader is then led to doubt whether anything Plato says in the *Phaedrus* is truly what he thinks, or is trustworthy.

However, with more thought, the reader may then realize that Plato is pointing out the frivolity and inadequacy of written words because he wants us to discover that it is what he leaves unsaid that is crucial to the message of the *Phaedrus*. To reach that realization, we have to first see that a text can be non-seriously written but have a serious intent or message. I suggest that Plato has written the *Phaedrus* in such a way that only the part of the audience that understands that Plato actually has a serious message will receive the message – namely, that the reader’s response is a more significant factor in determining whether a text is living or not. In order to comprehend this in the first place, the reader must engage in a philosophical question-and-answer with the *Phaedrus*, effectively turning it into a living text. Plato does not explicitly broadcast his message through Socrates because he wants readers to go through this philosophical process. In this way, Plato does not have to use explicit written words – which are inadequate for conveying reality – to lead his readers to truth. Instead, he can point us in a direction such that we can discover it by ourselves. Through the process of responding to a text, we learn that we have the ability and responsibility to bring a text to life, and then realize that we have already been doing so.

In conclusion, although Socrates neglects to note that the reader can play an active role in turning a text into a two-way dialogue instead of being merely the passive recipient of a one-way transmission of knowledge, I suggest that this is the actual message of the *Phaedrus*. I argue that Socrates’ neglect is a deliberate move on Plato’s part, as is Plato’s apparent indirect claim that the *Phaedrus* is not to be taken seriously by readers: Plato wants to limit his audience to those who are willing to devote time and effort to figuring out the actual intent of the *Phaedrus*. After figuring out Plato’s message that readers are responsible for breathing life into a text, these readers will then realize that to obtain this message in the first place, they have actually been breathing life into the *Phaedrus* all along. In this way, Plato does not have to use explicit written words, which are necessarily inadequate and imperfect for capturing reality, to give people some understanding about reality.

---

## Works Cited

Plato. "Phaedrus." Plato: Complete Works. Trans. Alexander Nehamas and Paul Woodruff. Ed. John M. Cooper. Indianapolis: Hackett, 1997.

----- "Ion." Plato: Complete Works. Trans. Paul Woodruff. Ed. John M. Cooper. Indianapolis: Hackett, 1997.

Lear, Gabriel Richardson. Personal discussion during office hours. 9 Dec. 2009.

Mosher, Hannah. Personal discussion during office hours. 7 Dec. 2009.



# The Third Man Argument, *Parmenides* 132a1-b2

IAN WELLS

*Cornell University*

**Abstract:** Over the last half-century, Plato's Third Man Argument [TMA] has received a surge of attention. The challenge which numerous critics have undertaken is to provide a viable interpretation of Plato's puzzling passage at *Parm.* 132a1-b2. The exegetical part of this paper attempts to bring together some of the most plausible interpretations offered to date, distilling the good moves from the bad. The not-so-exegetical part of this paper draws out the consequences of these most plausible interpretations. In particular, it considers the possibility, inspired in the first instance by a stage-functional interpretation of Plato's one-over-many principle, that Plato held a recursive theory of knowledge.

In Part I, I give a textual and logical analysis of the TMA. I try to formulate the TMA such that it validly generates a regress from consistent premises, while remaining faithful to the text. In Part II, I ask whether Plato is vulnerable to the TMA so conceived. I argue that some textual evidence suggests that he is not. In Part III, I assume for the sake of argument that Plato is vulnerable, and ask: (1) Is the conclusion of the TMA vicious – does it pose a problem for Plato? And (2), what are the consequences of the TMA (if it goes through) for Plato's claim that knowledge is possible?

## I. Textual and Logical Analysis of the TMA

Gill and Ryan's text (in Cooper 1997) reads:<sup>1</sup>

I suppose you think (a) each form is one on the following ground: (b) whenever some number of things seem to you to be large, perhaps there seems to be some one character, the same as you look at them all, and from that you conclude that (c) the large is one.

That's true, he said.

---

<sup>1</sup> A few notes: While I use Gill and Ryan's translation for now, I intend to discuss its relative merits, on each specific passage, as they arise. (We will be inclined to reject some of these translations in favor of alternatives.) Also, I am deploying the same notational referencing scheme as Fine, also for 'ease of reference' and clarity (Fine 1993, 204; Fine 2009, class handout).

---

(d) What about the large itself and the other large things? If you look at them all in the same way with the mind's eye, again (e) won't some one thing appear large, by which all these appear large?

It seems so.

So another form of large will make its appearance, which has emerged alongside largeness itself and the things that partake of it, and (f) in turn another over all these, by which all of them will be large. (g) Each of your forms will no longer be one, but unlimited in multitude (*Parm.* 132a1-b2).

Let's begin with (a).<sup>2</sup>

a. The Uniqueness Assumption

Context tells us that (a) is the target of the TMA – i.e. the Platonic assumption which Parmenides aims to disprove. Call it the Uniqueness Assumption (UA). Let's suppose, at least initially, that (UA) is just as it appears in the G-R translation:<sup>3</sup>

(UA) Each form is one.

This is mysterious. Does (UA) amount to the claim 'each *singular* form is *singular*'? If so, it would be trivially true, in the same way the proposition 'every white fence is white' is trivially true.<sup>4</sup> But surely, Parmenides wouldn't object to such a truth (as it is scarcely objectionable); nor would Plato include it in his theory of forms (as it adds nothing of substance to the theory).<sup>5</sup> Rather, there

---

<sup>2</sup> The order of my analysis deviates importantly from that of Vlastos (*Vlastos I*, 320). He jumps right into an analysis of the explicit premise (b), and the conclusion (d)-(g), without first establishing the target of the TMA – that is, (a) – which is to determine much of how we interpret the passage as a whole. I think this is precisely where Vlastos gets into trouble. We will see how he ends up building the Uniqueness Assumption at (a) into his formulation of (b). This is problematic, as will be shown (see Fine 1993, 208; and my Section I, Part b). I take this fact as a warning: We must first fully clarify (a) before trying to formulate the rest of the argument.

<sup>3</sup> This is also Fine's translation of (a) (Fine 1993, 204; Fine 2009, class handout).

<sup>4</sup> The triviality becomes explicit if we strip this proposition down to its formal structure: ' $\forall x((Px \wedge Qx) \supset Px)$ ' or, even more deliberately, ' $\forall x(Px \supset Px)$ .' In model-theoretic terms, any model (any valuation of the predicates involved; any universe) would make these propositions true; the meaning of the logical connectives alone guarantees their truth. But Plato clearly wants his theory of forms to take some stand on the state of the universe and on how the words 'form' and 'one' are to operate.

<sup>5</sup> Here my reading is indebted to Cohen, who also dispenses with such a translation on the same grounds of "triviality" (Cohen 451). However, I don't think we *need* to read (UA) in this way. Cohen seems to be presupposing the presence of a certain conversational implicature attaching to the term 'each', such that: 'each  $x$ '  $\rightarrow$  'each *one*  $x$ ' or 'each *single*  $x$ '. But this is adding extra semantic content where it may not necessarily occur. Perhaps people do in fact interpret natural language in such a way (this is a matter for semantic or linguistic theory), but the fact remains that we need not read (UA) in this way. I propose a different, nontrivial reading in what follows. Thus, I

---

must be something else going on here. Perhaps Plato means: ‘for every  $x$ , if  $x$  is a form, then  $x$  is one (in number)’.

Here, we no longer build ‘being one’ into the protasis; this omission appears, at least *prima facie*, to alleviate any temptation toward reading (a) in the trivial sense, since (UA) now looks like a statement of the form ‘ $\forall x(Px \supset Qx)$ ’, e.g. ‘every fence is white’. But this reading is also peculiar. Does ‘one (in number)’ mean that  $x$  is one *form*, in the same way we say that this book is ‘one book’? That wouldn’t provide us with much information about forms at all, besides perhaps that there are not mini-forms ‘inside’ each form, in some immaterial sense of ‘inside’; but this doesn’t seem to be the Platonic principle at stake in the TMA.<sup>6</sup> Such an interpretation, of course, could also reintroduce the trivial reading, if we understand ‘each’ to carry with it an implicit ‘one’, such that (UA) states: ‘each one form is one form’. So this won’t do. What if we understand ‘ $x$  is one (in number)’ to mean, more precisely, ‘ $x$  is the one form corresponding to the sorts of things it is a form of’? For example, ‘the form of largeness is the one form corresponding to large things’.<sup>7</sup> This is getting better, but it’s still a far cry from our initial translation (a).

Notice what all of these readings of (UA) have in common: They attempt to answer the question ‘How many forms are there?’ by saying that every form has the *property* (‘ $Q$ ’) of ‘being one’ (however that property is to be understood). And they *must* answer the question in this way, due to their formal structure: ‘ $\forall x(Px \supset Qx)$ ’. ‘ $P$ ’ is just the property of being a form; oneness only comes in at ‘ $Q$ ’, as a further property that all forms are to possess. We can massage this formula to mean something approaching what we want it to mean, but only by squeezing the property of ‘being the one form corresponding to the sorts of things  $x$  is a form of’ in as ‘ $Q$ ’. And besides being painfully unclear, such a property presents a more pressing problem: It seems to include a hidden

---

don’t think we can dispense with (UA) altogether, based on the triviality objection; but the objection does give us good reason to reject one particular interpretation of (UA). It also gives us good reason to try to avoid even the possibility of interpreting (UA) in such a way. This fact informs my decision, later, to reject Gill/Fine’s translation of (a). I do, however, have another, more forceful reason for this decision.

<sup>6</sup> As we shall see, the conclusion of the TMA does not threaten (a) the ‘oneness’ or ‘wholeness’ of each *single form*; rather, (b) it posits *multiple* (here, infinitely many) *forms* for each single predicate (although it is debatable whether every predicate must have a form, and, consequently, how we should read this latter claim). This difference between (a) and (b) can be made explicit in quantificational terms: (a) universally quantifies over forms; (b) universally quantifies over predicates. I talk about this in what follows.

<sup>7</sup> Note that I have introduced a ‘corresponding to’ relation here which I have not yet discussed. Precisely what Plato might take it to mean will become more evident as our discussion progresses. For now, however I do not take a stand on how we should read this relation.

---

existence claim, snuck in under the guise of ‘the’ in ‘being *the* one form.’ But nothing is existentially quantified in Gill/Fine’s translation. So this reading, which we’ve found to be the only one even close to capturing what we want (UA) to mean, departs altogether from the logical structure of Gill/Fine’s translation. Furthermore, even if we were to suppose that there is no hidden existence claim, and that we could get the point across simply by ascribing a *property* to forms, we would want to ask: Is number (e.g. ‘being one’) actually a property of things? Or, rather, does Plato take it to be? Perhaps he does, and perhaps this line of interpretation is okay (despite the flaws I have identified). But before making our decision, let’s look at an alternative translation of (a) which avoids all the aforementioned problems, and, in turn, motivates an alternative formulation of (UA).

Cornford translates (a): “there is one form in each case” (Cornford, 87). Cornford’s translation conflicts with Gill/Fine’s on two accounts. The first disagreement is over the scope of ‘each’: Gill/Fine take it to universally quantify over forms; Cornford takes it to universally quantify over ‘cases’. (Resolution of this disagreement demands an answer to the following question: Is Plato’s (UA) meant to obtain for all forms, or for all ‘cases’, whatever that means? See Note 12.) The second, and more glaring, disparity between the translations is this: Cornford includes a nested existential quantifier within the scope of his universal, whereas Gill/Fine omit the existential quantifier altogether.<sup>8</sup> Two points about this: First, while it is possible to interpret (UA) as trivially true on the Gill/Fine translation (the main logical constant being a universal quantifier), we find Cornford resistant to such an objection, since it makes an existence claim).<sup>9</sup> Second, Cornford’s (a) answers the question ‘How many forms are there?’ quite differently than Gill/Fine’s (a). Cornford’s makes no use of the property ‘being

---

<sup>8</sup> Of course, Cornford’s (a) could be translated differently than I’ve done (with the existential nested within the universal): the order of quantifiers could be switched with a little logical work. However, I take my translation as the most natural, since the alternative would involve a conjunction (which we don’t find in the English). The main point is that Cornford’s (a) includes two instances of quantification, including an existence claim, whereas Gill/Fine’s does not.

<sup>9</sup> Any claim to existence must be nontrivial. This is because we can always imagine a universe in which nothing exists. So, given any existence claim (any existentially quantified proposition), there will be at least one possible world in which that claim is made false (i.e. the world in which nothing exists). And this holds no matter how trivial the thing being claimed to exist turns out to be (e.g. ‘there is some *x* such that *x* is a duck iff *x* is a duck’). An objector might wonder whether such an ‘empty’ world is possible, or whether we can actually imagine such a world. This would be an interesting discussion but is outside of the scope of this paper. Insofar as logic is concerned, existence claims are nontrivial.

one'; forms don't *have* this property. Rather, there just *is* one form for every case.<sup>10</sup> This seems to capture the sense of the hidden existential claim in our best interpretation of Gill/Fine's (a). So, to capture this sense, and to avoid all the aforementioned problems that Gill/Fine's (a) opened, I propose to take the existential route suggested by Cornford.<sup>11</sup> (For another reason why Cornford's translation is more appealing, see Note 12.) So let's try

---

<sup>10</sup> Fine takes this quality of Cornford's translation to be a defect: "[Cornford's] translation suggests that Plato is explaining why he believes that forms exist, whereas my translation... suggests that he is explaining why there is just one form for any predicate that has a form" (Fine, fn. 8). To be sure, I don't think Fine's (a) suggests that "*there is* just one form for any predicate" at all (though her (UA) does). If it did, her argument here would be self-defeating, since it targets precisely those translations which posit the existence of forms. Rather, as we've seen, the only quantifier Fine's (a) includes is the universal. So, more accurately, Fine's (a) suggests that all forms *have the property* of 'being one.' That said, Fine's point is a good one. Embedded in Cornford's translation is a claim that *there are forms*. Fine's translation merely says, 'if there are forms at all, then these forms are of *such a sort*'. And this, Fine thinks, better captures the sense of the (UA): Plato doesn't mean for the (UA) to be an argument *for forms*; rather, it's an argument for (or, a claim to) a certain fact *about forms*. I agree that this "argumentative context" favors Fine's translation.

However, Fine is presupposing that we *can* express facts about the number of things without making existence claims about those things – e.g. that 'being one' or 'being two' can be expressed just like any other property. I think the issue between the translations boils down to a rather difficult philosophical question: How are we to understand statements of number? Fine's translation (again, not her (UA)) presupposes the Fregean and perhaps Platonic view that numbers are properties of objects. Cornford's presupposes the view that numbers arise out of existence claims, together with statements of identity: We arrive at the number 'one' by claiming 'there is an x' (numerical modifiers like 'exactly' would require the addition of universal quantifiers and identity; no numerical predicate necessary); similarly, we arrive at the number 'two' by claiming 'there is an x and there is a y, and x does not equal y.' The downside of this way of framing the concept of number is that it introduces into Plato's (UA) a claim about the existence of forms (making it seem like he is trying to explain why they exist). The upside is that it does *not* introduce into Plato's (UA) any claim about how we are to understand statements of number, as Fine's (a) does. Such a claim, after all, would be equally unfitting to the "argumentative context" of the (UA). (The TMA cares little of such things.) So each translation carries with it theoretical baggage seemingly external to the TMA, none of which we really want Plato to have to commit to, or explain, in defending his (UA). (I end up choosing Cornford's (a) for reasons independent of this debate, which I remain neutral on; my point here is to level the playing field after Fine's charge, and to demonstrate how there may be more deeply-entrenched Platonic, or otherwise philosophical, issues, in relation to the question of number, at play which have been swept under the carpet.)

<sup>11</sup> This is, strangely, the route Fine ends up taking, adopting the following as (UA): "There is exactly one form corresponding to every predicate that has a form" (Fine 1993, 205). I don't know how Fine's translation of (a) leads to such a formulation of (UA) – in fact, this (UA) looks more like Cornford's translation of (a) (which Fine of course rejects) than her initial translation. This is because in Fine's (UA) 'forms' are quantified *existentially* ('there is', as in Cornford's (a)), while

---

(UA<sub>1</sub>) There is one form in each case.

This is better, but still won't do. The existential quantifier here only tells us that there is *at least* one form in each case. But what Parmenides is arguing against, in (b)-(g), is the Platonic notion that there is *just* one form in each case. Without yet dipping into these middle passages at (b)-(g), we can conclude by rough approximation that they aim to show that where Plato thought there was just one form, there are actually "unlimited" forms. (Here I've presupposed, for the sake of argument, a reading of (g) which we have not yet defended or accepted.) In any event, if (UA<sub>1</sub>) allows for the possibility of multiple forms of F, then it must not be the Platonic assumption Parmenides wishes to reject, nor the one Plato holds. (This will become clearer later.). Let's limit it, then:

(UA<sub>2</sub>) There is exactly one form in each case.

But this is ambiguous: How should we understand the phrase 'in each case'? Again, let's peek at the middle passages, (b)-(g), to get the gist of the TMA. Here we find an argument that takes as its subject one *particular* quality/character – that of 'being large' – and shows how if there is even just one form of large, then there must be infinitely many. So if (UA<sub>2</sub>) is to be the general Platonic principle under attack, then the 'cases' it refers to must be those of individual qualities or characters, e.g. 'being large.' The TMA, then, takes an instantiation of one of these 'cases' – the predicate 'being large' – and shows how it violates the principle. Since this principle is supposed to apply *in general* (i.e. since it is *universally* quantified over 'cases'/predicates/qualities), such a demonstration of *just one* predicate's violation of the principle will suffice in falsifying that principle. Therefore, we had better clarify:

(UA<sub>3</sub>) There is exactly one form for (alt: corresponding to) every predicate that has a form.<sup>12</sup> This is the version of (a) that I will accept as sufficiently clear.<sup>13</sup> We can move on now to the actual argument.

---

in her (a) there is no existential quantification whatsoever. Correlatively, Fine's (UA) abandons the sense of her (a) which took 'being one' to be a property.

<sup>12</sup> I have not justified the inclusion of the last condition on every predicate: 'that has a form'. This condition is suggested by Fine (Fine 1993, 205; Fine 2009, class handout; against Peterson 451). She argues that Plato is not committed to the assumption that every linguistic predicate has a corresponding form, as Aristotle thinks he is (Fine 1993, ch. 8). Thus, Plato's (UA<sub>3</sub>) should be restricted only to those predicates ('cases') that have a corresponding form, so that the TMA poses an eminent threat and cannot be so easily dispensed with on the grounds of Plato's lack of commitment to an unconditioned version of (UA<sub>3</sub>). This seems accurate to me, though I will not have time in this paper to investigate it thoroughly. But one point is worth mentioning: Note how easy it would be to build Fine's condition into Cornford's translation. Such ease is engendered by the fact that Cornford's (a) makes the (UA) out to be a principle concerning all *predicates*, but not

---

### b. Take 1: Explicit Premise, Implicit Premises

Recall where we're at: Parmenides has sketched the Uniqueness Assumption (a); we've tried to clarify what it might mean (UA<sub>3</sub>). Parmenides then inquires into Plato's "ground" for accepting (UA<sub>3</sub>) (*Parm.* 132a1). This "ground" is presented at (b), and it forms the singleton premise set of the TMA. The conclusion of the TMA, however, is not what Plato might expect. The premise that he takes to affirm his general principle (UA<sub>3</sub>) actually entails something much different. Precisely what this premise entails will become evident in what follows, although we can already get an idea of what it *should* entail if the TMA is to disprove (UA<sub>3</sub>). The proper negate of (UA<sub>3</sub>) is:

(¬UA<sub>3</sub>) It is not the case that there is exactly one form for (alt: corresponding to) every predicate.

Or, equivalently,

(¬UA<sub>3</sub>) There is some predicate such that there is not just one form for (alt: corresponding to) it.

---

necessarily all forms. So all we need to do is build the conjunct 'has a form' into the protasis of the conditional:

(a)  $\forall x(Px \supset \exists y(Fy \wedge Ryx))$  (omitting the limitation to *exactly one* 'y', for brevity)

(UA<sub>3</sub>)  $\forall x((Px \wedge Qx) \supset \exists y(Fy \wedge Ryx))$  (ditto)

(Where 'P' = the predicate for 'being a case/predicate', 'F' = the predicate for 'being a form', and 'R' = the two-place relation for 'corresponding to'.)

The easy move from (a) to (UA<sub>3</sub>) is another reason why I favor Cornford's translation over Gill/Fine's; it's also another reason why such a translation seems to better suit Fine's *own* valid formulation of the Uniqueness Assumption (and her arguments for such a formulation (Fine fn. 9, ch. 8)). How would we insert the stricture 'y has a form' (obtaining for *every* predicate y) into Gill/Fine's translation of (a), when we aren't quantifying over predicates to begin with? Perhaps this could be done, but not without significant logical work. In any event, we see how Cornford's translation of the Greek gets us closer to the logical meaning of the UA we wish to exemplify: The upshot is that we don't have to foist upon Plato a Uniqueness Assumption that bears little semblance to what he actually says at (a). Lastly, we can now answer the previously mentioned question: Is Plato's (UA) a principle obtaining for all forms, or for all 'cases'/predicates? As I've argued, it seems best to take the (UA) as a principle universally quantified over predicates (captured by Cornford's (a)), not forms (captured by Gill/Fine's (a)). We can finally close our rather lengthy discussion of how to translate (a); I harp on the issue only because it will help to nail down precisely what Platonic principle serves as the backdrop – the target – of the TMA. For otherwise, how would we know what Parmenides is arguing against? How would we know whether his explicit premises actually tempt but do not imply the (UA)?

<sup>13</sup> There still seems to be some vagueness. My final version, (UA<sub>3</sub>), introduces a new relation – that of 'being for' or 'corresponding to.' We have not yet talked about this 'correspondence' relation between forms and predicates, and I don't intend to at this juncture. Hopefully it will become clearer as we move on.

---

So all the TMA needs to show is that there is some predicate which has more than one form corresponding to it. This is precisely what it concludes (and more), taking as its example the predicate ‘large’. Let’s see how it proceeds towards this conclusion.

The single explicit premise in support of (UA<sub>3</sub>) is (b). It is followed immediately by the conclusion of the TMA, which comes, in parts, in (d)-(g). There are several ways of reading the initial explicit premise and the subsequent conclusion. In keeping with the literature, I will begin first with Vlastos’ interpretation. Vlastos formulates the premise as follows:

- (A1) If a number of things *a, b, c* are all F, there must be a single Form, F-ness, in virtue of which we apprehend *a, b, c* as all F (Vlastos I, 232-233).

From this premise, Parmenides infers the conclusion at (d)-(f), which Vlastos also formulates:

- (A2) If a number of things *a, b, c* and F-ness are all F, there must be another Form, F-ness<sub>1</sub>, in virtue of which we apprehend *a, b, c*, and F-ness as all F (Vlastos I, *ibid*).

Vlastos accurately observes that (A2) is not a formal consequence of (A1) alone. We can see this quite easily if we strip (A1) and (A2) down to their essentials:

- (A1) If some things are F, then they are F in virtue of F-ness.

- (A2) If some things are F, then they are F in virtue of F-ness<sub>1</sub>.

These look like two entirely separate claims (‘if P then Q’, ‘if P then U’). And there is no way, just by considering (A1), to deduce (A2). Thus, Vlastos concludes that Plato had in mind two other implicit premises, which enable the logical jump from (A1) to (A2) [Vlastos I, 236]:

- (SPV) Any form can be predicated of itself. Largeness is itself large. F-ness is itself F.

- (NIV) If anything has a certain character, it cannot be identical with the form in virtue of which we apprehend that character. If *x* is F, *x* cannot be identical with F-ness.

Do these implicit premises help make the TMA valid? Let’s see. Given a set of ‘F-things,’<sup>14</sup> we conclude by (A1) that they are F in virtue of F-ness. By (SPV), this F-ness is itself F. Taking the set of F-things together with F-ness, we conclude by (A1) again that they are all F in virtue of a second F-ness, call it F-

---

<sup>14</sup> I will assume, to begin with, the existence of F things as given, that is, that there really is large stuff; that we ‘see’ it every day and that we can, consequently, talk about it. This will be an assumption built into most of our discussion of the TMA – and I take it that when Plato mentions “some number of things” at the outset of the TMA passage, he too makes this assumption.



ness<sub>1</sub>. By (NIV), F-ness is not identical with F-ness<sub>1</sub>. Thus, there are now two forms for (corresponding to) F. (Note: This is enough to prove (A2), and it is already sufficient for a rejection of (UA<sub>3</sub>).) But, by (SPV), F-ness<sub>1</sub> is F. Now consider the set of F-things, F-ness, and F-ness<sub>1</sub>. By (A1) again, we arrive at F-ness<sub>2</sub>, ... *ad infinitum*.

This is one way to arrive at (A2) and generate a regress using Vlastos' premises together with (A1). But there is another way that doesn't involve (A1) at all. This is relieving because (SPV) and (NIV) are inconsistent: We can derive a contradiction from them alone.<sup>15</sup> Here's how: (NIV) is in effect a universally quantified statement of the form, ' $\forall x(Fx \supset x \neq F\text{-ness})$ '. (SPV) is a fact that says that some individual constant (F-ness) has the property F. Thus, F(F-ness) is true. By informal universal elimination on (NIV), we get ' $F(F\text{-ness}) \supset F\text{-ness} \neq F\text{-ness}$ '. The protasis is true, so we can conclude that ' $F\text{-ness} \neq F\text{-ness}$ '. But everything is equal to itself. Hence, absurdity; hence, we may conclude whatever we'd like (say, (A2)). We see, then, how (SPV) and (NIV) lead to a contradiction all by themselves and are thus inconsistent.

We've shown that Vlastos' initial formulations of (SPV) and (NIV) generate a regress, and they do so validly. But they do so in virtue of their inconsistency. As Cohen noted, if the premises of a derivation are both inconsistent and necessary (not just sufficient) for some conclusion, then that conclusion must be a contradiction (Cohen 451). Vlastos knew (SPV) and (NIV) to be inconsistent, yet he also thought them necessary: "we need... the Self-Predication Assumption," "we need... the Non-Identity Assumption" (Vlastos I, 324-5). Thus, Vlastos is committed to saying that the conclusion of the TMA is a contradiction. But is it? Cohen thinks not, and for good reason. Of course, before demonstrating this, we'll need a working formulation of the conclusion of the TMA.

The Gill-Ryan conclusion at (g) states: "Each of your forms will no longer be one, but unlimited in multitude" (*Parm.* 132b2). Fine translates similarly (Fine, 204).<sup>16</sup> This translation is just as mysterious as the Gill/Fine translation of (a). Again we could ask, with Cohen, whether 'each' carries with it a tacit 'single', such that (g) amounts to: 'each single form will no longer be

<sup>15</sup> Vlastos noticed this, but Peter Geach formalized it (Geach 1956, 72-82).

<sup>16</sup> We find a different translation on the class handout, which I find more appealing (Fine 2009): "So there will no longer be one of each of these forms for you, but infinitely many." This statement includes a negation of an existential claim, and it lacks any universal quantification over forms, which makes it look more like the proper negate of Cornford's translation of (a) and less like a negation of Gill/Fine's (a). I therefore prefer this translation.

single'. In this case, (g) looks like a contradiction, and perhaps Vlastos is right to say that his inconsistent premise set is necessary. But then his choice of (SPV) and (NIV) becomes arbitrary: for, as Cohen notes, any two inconsistent premises "would do just as well" in generating the contradiction at (g) (Cohen 451). So, in that case, (SPV) and (NIV) in particular would be merely sufficient, not necessary, for the argument. Let's try another case. Suppose we need not build a tacit 'single' into (g). In that case, (g) appears consistent, and Vlastos' implicit premises must not be necessary. Let's try, lastly, a third case. Suppose that the Gill/Fine translation is misleading and that we should rather adopt a translation of (g) more along the lines of Cornford's translation of (a). This is the reading of (g) I favor.<sup>17</sup> Thus, I suggest the following reading of (g), which we find in Cohen (ibid):

(g) "there will no longer be one Form for you in each case, but infinitely many"<sup>18</sup>  
 This reading of (g) certainly does not make the conclusion seem to be a contradiction. So, in all three cases, we've seen that neither (SPV) and (NIV), nor any other *particular* set of inconsistent premises, will be necessary for deriving the conclusion at (g). If they aren't necessary, then why assume Plato held them? We *need* to supplement the explicit premise in order for the TMA to pose a formidable threat to Plato – but we want to do so responsibly. Saying that the TMA works because Plato implicitly thought ' $A \wedge \neg A$ ' (and that anything follows from that) will not suffice,; unless, of course, it's the only option. Our challenge, then, is this: To seek anew a *consistent* premise set that validly entails a regress when conjoined with (A1).

### c. Take 2: Implicit Premises, Explicit Premise

Both Sellars 1955 and Fine 1993 (among others) answer our challenge. Since there are a few problems with Sellars' formulation of the implicit premises, let's skip ahead to Fine's.<sup>19</sup>

---

<sup>17</sup> My argument here is parasitic on my argument for Cornford's translation of (a). This is because, as I suggest, the conclusion of the TMA aims to *deny* (a) (i.e. the Uniqueness Assumption), and should thus stand in (approximate) antithetical relation to it. (I include 'approximate' since, as we know, the conclusion overshoots its goal).

<sup>18</sup> Of course, our (g) is not, strictly speaking, the proper negate of the Uniqueness Assumption (UA<sub>3</sub>) that we developed out of Cornford's (a). But it may be. That negation, remember, was just a *goal* for the TMA. The conclusion only needs to *meet* that goal, and it does. In fact, it overshoots: All that was needed was to demonstrate one instance of there being more than one form corresponding to some predicate in order to deny (UA<sub>3</sub>). This reading of (g) also stays faithful to the logical apparatuses at work in (a): Forms are quantified existentially; cases are quantified universally.

---

(SPF) Any form of F is itself F.

(NIF) Nothing is F in virtue of itself.<sup>20</sup>

Whereas (SPV) said that every form of F can be predicated of itself (i.e. every form of F is F *in virtue of itself*), (SPF) omits this explanatory addendum altogether. (NIF) captures more succinctly Vlastos' original. Further, (SPF) and (NIF) are consistent. (See Item 1 in the Appendix.)

But do they generate a regress? Not on their own. We would need to add a formulation of the explicit premise at (b). Sellars formulated it as such:<sup>21</sup>

(bS) If a number of things *a, b, c* are all F, there must be a single Form, F-ness, in virtue of which they are all F (Sellars, 409).

So conceived, (bS) validly generates a regress when taken in conjunction with (SPF) and (NIF). Here's why: Given some F-things (*a, b, c*) as usual, (bS) tells us that these things are all F in virtue of a single form, F-ness. (NIV) tells us that none of these F things is identical to F-ness. (Why? Assume they were: Since they are F in virtue of F-ness, then there is something that is F in virtue of itself; this contradicts (NIF). Discharging the initial assumption, we conclude that our F things are not identical to F-ness.) (SPF) tells us that this F-ness is itself F. So we now have a new member of the set of F-things. Consider this new member together with the original Fs; apply (bS) to this composite set; we'll get a similarly new F-ness, identical neither to the original F-ness nor to the original Fs. And so on, *ad infinitum*.

Since Fine's implicit premises are consistent, and they validly imply a regress in conjunction with (bS), this new formulation of the TMA looks much better than Vlastos'. But there is a problem. As Vlastos noted in his second TMA paper, Sellars' (bS) is not faithful to the text (cf. Vlastos II, 293). Vlastos has

---

<sup>19</sup> Cf. Cohen fn. 14. We must, however, credit Sellars with locating and remedying the problem that generated inconsistency in Vlastos' original premises. This problem turns on how we are to read 'F-ness.' Sellars argues that 'F-ness' should be considered a representative variable: It stands for (i.e. represents) a set of variables. If F-ness is a variable (of this sort) it must be bound by a quantifier, or else it will occur free (as Vlastos' premises have it). Thus, Sellars adds quantifiers to SP and NI. This discovery surely enabled Fine's formulation. However, Fine's final version is simpler and less problematic, so I will use it.

<sup>20</sup> Fine notes that (NI) contains a claim to *strong* non-identity (Fine 207). That is, it's not just the case that sensibles are not F in virtue of themselves. *Nothing* (either sensible or immaterial, as forms) is F in virtue of itself. This is a point Fine attributes to Vlastos (Vlastos I, p. 238), though he used the term 'particulars' in place of Fine's 'sensible.' Surely, it's easy to see how the TMA needs strong (NI), and would not generate a regress with merely weak (NI), applicable only to sensibles.

<sup>21</sup> See also Strang 1963 for a similar reading.

several supporting arguments for this claim; I'll mention only the most compelling one. Recall that the explicit premise is not only meant to validly generate a regress (to Plato's dismay), but also to *tempt* Plato into believing (UA<sub>3</sub>). Does (bS) *seem* to "ground" the claim, 'There is exactly one form for every predicate that has a form'? Vlastos thinks not. To say that there 'must be a single Form' doesn't rule out the possibility that there could be more than one form.<sup>22</sup> This condition of having a 'single Form' could be satisfied even if there are multiple forms. As Vlastos noted, what we need to include in (b) is a restriction to '*exactly* one form', in order for (b) to appear compelling as a "ground" for (UA<sub>3</sub>). Thus Vlastos offers his own, more faithful, reading of (b):

(bV) If a number of entities are all F, there must be exactly one Form corresponding to the character F; and each of those entities is F by virtue of participating in that Form.<sup>23</sup>

Now this is textually appropriate. But the problem with (bV) is that it combines with (SPF) and (NIF) to form an inconsistent set of claims (see Item 2 in the Appendix). So if we accept (bV), we have failed in our challenge to validly generate a regress with a consistent premise set.<sup>24</sup> But should we accept (bV)? There seems to be a problem. As Fine points out, (bV) doesn't serve its proper role in the TMA, and it commits the same crime of infidelity to the text as (bS). As we've seen, the explicit premise is supposed to *tempt* Plato to think (UA<sub>3</sub>) obtains; but, of course, it should not *actually entail* (UA<sub>3</sub>). But (bV) builds into the TMA's explicit premise the distinctive claim of (UA<sub>3</sub>) – namely, the 'uniqueness' claim, the claim that there is '*exactly* one Form' – such that (b) entails (UA<sub>3</sub>). It simply *states* (UA<sub>3</sub>). Thus, Vlastos goes "too far in the other direction" (Fine, 209). He rightly concludes that Sellars' (b) is not tempting

<sup>22</sup> To be fair to Sellars, I think both Cohen and Fine paraphrase him incorrectly. His actual words in the paper are: 'a single Form' – not, as Cohen has it, '*an F-ness*', nor, as Fine has it, '*at least one form*'. Sellars' original formulation looks stronger (more like '*exactly one*') than either of these paraphrases. Nevertheless, it is not unwarranted to demand stricter clarity and to limit (bS) to *exactly one form*.

<sup>23</sup> This is Cohen's adaptation of Vlastos' (b) (Cohen 455; Vlastos 290). I use it here because it is syntactically consistent with the terminology we've been using throughout this paper (whereas Vlastos' is not), and because Cohen's is an accurate paraphrasing of Vlastos'. In his paper, Vlastos uses the term 'unique' instead of 'exactly.' But, given his conclusions toward the end of the paper, it is clear that we can understand 'unique' to mean 'exactly' (Vlastos 296).

<sup>24</sup> Vlastos knew his (b) would mark the end of the struggle, as he thought it to be the only textually faithful rendering of (b) that validly generated a regress. Since it made the premise set inconsistent, he concluded that Plato must have formulated the TMA as such. Vlastos' conclusion is not that Plato was trying to pass a bad argument as a good one, or that he was trying to deceive us, but that the TMA is "a record" of his "honest perplexity" (TMA I 254).

---

enough, but wrongly concludes that (b) should include the very notion (of uniqueness, exactitude) that it aims to disarm (by Parmenides' lights) and "ground" (by Plato's initial lights). Since (bV) is problematic, perhaps it is not our only option, and perhaps we can resume our challenge. This time, though, we must pursue a bipartite goal: (i) formulate a *consistent* premise set that validly entails a regress when conjoined with (b), and (ii) formulate (b) such that it remains faithful to the text (i.e. tempts Plato to, but does not imply, (UA<sub>3</sub>)).

One alternative to (b) reads:

(bW) If a number of *sensibles* are all F, there must be exactly one Form corresponding to the character F; and each of those entities is F by virtue of participating in that Form<sup>25</sup>

This certainly doesn't imply (UA<sub>3</sub>), since it limits the claim to cover only sensibles having some quality F (whereas (UA<sub>3</sub>) holds universally). In this way, it is better than (bV). It also includes the numerical modifier 'exactly', which could tempt Plato to conclude (UA<sub>3</sub>). In this way, it is better than (bS), which made no mention of 'exactly one' form over the Fs but left open the possibility of multiple forms over the Fs. However, (bW) is clearly no good, since it doesn't combine with (SPF) and (NIF) to generate a regress. This fact is obvious: The cogs of the TMA only get going once forms, as well as the initial stock of F-sensibles, are thrown back into (b). If (b) can't process a bundle of F-objects that includes one or more (non-sensible) F-forms, then it won't spit out a fresh form; hence, the argument stops in its tracks. Of course, (bW) is not our only option; Cohen 1971 and Fine 1993 offer an alternative reading of (b) that succeeds where the rest have failed.

#### d. Final Take: A Stage-functional Explicit Premise<sup>26</sup>

---

<sup>25</sup> This is my own formulation, though it is similar to Fine's and Cohen's (Fine 210; Cohen 460). With hope of retaining some syntactic continuity in this paper, I simply added the condition of 'being sensible' to the F things in Vlastos' (bV), instead of adopting another critic's version. One note on (bW): It is *weak* in the sense that it only applies to sensibles, but not forms. Whether or not this is an "unobjectionable" Platonic truth – whether Plato clearly holds this weak notion to be true – is debatable (Cohen thinks so (Cohen, 460); Fine thinks not (Fine, n. 35)). Another note: Fine introduces the notion of a set here, at her version of (bW); that's certainly possible to do. It also motivates us, in seeing the problem with (bW), to look for a (b) "that posits exactly one form for (and only for) every set relevantly like the set that consists of all and only large sensibles" (Fine 210). This notion of 'relevance' is crucial to Cohen's formulation.

<sup>26</sup> Due to the density of this formulation, I will have to omit proofs of all the facts needed to fully justify Cohen's formulation. For details, cf. Cohen 1972. But the point I want to take from Cohen's formulation is that it deploys a stage-functional scheme, whereby we can speak of sets at specific stages. This notion, I think, fixes many of the problems that we've seen the other interpretations of (b) to open. It also informs much of my discussion in the last part of this paper.

---

In all the formulations of (b) that we have considered, we have talked about sets of things having some property/character F. To begin with, there are the F-things – sensibles – that we are given (e.g. we see large trees). Informally, the basic idea is that premise (b), in some way or another, gives us a form of F in addition to these F-things, which is itself F. This form, in turn, becomes a *member* of the original set of F-things... and so on. Cohen tries to formalize our talk of sets. Instead of quantifying just over F-things, he suggests we quantify over sets of F-things. But this is not enough.<sup>27</sup> We also need to introduce the notion of *stages*.<sup>28</sup> He wants us to speak of objects as being *at* a certain stage. Thus, we introduce ‘stage’ into our vocabulary set as a function constant. An understanding of how this function interacts with objects – of what it means for an object to be ‘at stage x’ – will emerge by observing the following conditions:

- (1) Let an object be at stage 0 iff it is a sensible object.
  - (2) Let an object be at stage 1 if
    - (i) all of its participants are sensibles (i.e. objects at stage 0), and
    - (ii) all sensibles (i.e. objects at stage 0) participate in it.
  - (3) Let an object be at stage 2 if
    - (i) all of its participants are either
      - (a) sensibles (i.e. objects at stage 0), or
      - (b) objects at stage 1
    - (ii) all objects at stage 0 and stage 1 participate in it.
- ... and so on.<sup>29</sup>

We see already how this structural scheme seems to capture the sense in which we have been talking about forms and sensibles, and the participation relation (elsewhere, ‘F in virtue of’). Now, since objects are members of sets, the ‘stage’

---

<sup>27</sup> Simply inserting the term ‘set’ into all our previous tries at (b) will not help much: Cohen demonstrates this during a long process of trial and error (Cohen 459-461). The notion of stagehood, rather than sethood, becomes the distinguishing mark of Cohen’s injunction – the point which gets the TMA off the ground, and in a way that seems faithful to the text. This is why I title my section (d) ‘Cohen’s Stage-functional Explicit Premise’ instead of ‘Cohen’s Set-theoretic Explicit Premise’.

<sup>28</sup> Fine suggests that this introduction of a new theoretical term may be very reasonable in light of the text. Plato himself talks of degrees of reality: “Plato plainly distinguishes between different levels of reality; in his view, forms, for example, are ‘more real’ or have ‘more being’ than sensibles do” (Fine 210).

<sup>29</sup> We could formalize this ‘and so on’ with a general definition: ‘Let an object be at stage n (where  $n \geq 1$ ) if (a) all of its participants are of stage  $n - 1$  or lower, and (b) all objects at stage  $n - 1$  or lower participate in it’ (where  $n \in \text{Natural numbers}$ ). See also, Cohen 461.

function also covers sets. So, given the above scheme, Cohen defines the stage of a *set* of objects as the stage of its highest-stage member; and a *maximal set* as the set that contains every object at every stage equal to or less than the stage of its highest-stage member (Cohen 462; see also Fine 210 for similar formulations). Of course, the stage of a maximal set (like any set) is the same as the stage of its highest-stage member. Now, these are the general definitions. If we are dealing with sets of F-things, then we'll have to tack this condition of being F on to our definitions: (i) the stage of a set of F-objects is the stage of its highest-stage member; (ii) a maximal set of F objects is the set that contains every F-object at every stage equal to or less than the stage of its highest-stage member. Let's look at an example to clarify. Take, for instance, all sensible large things. The set of all sensible large objects is at stage 0, since the stage of its highest-stage member is 0 (since all sensibles are at stage 0). This set is also a maximal set of large objects at stage 0, since it includes every large object at stage 0 (the stage of its highest-stage member) and every large object below stage 0 (of which there are none). This is the general picture. But a regress doesn't get going unless we introduce a one-over-many principle, which will tell us what sorts of objects are at the higher stages, and how many. We now have the tools to do so (note:  $N$  = the set of all natural numbers):

(OM) For any  $n \in N$ , and for any maximal set of Fs at stage  $n$ , there is exactly one form of F at stage  $n + 1$  participated in by all and only the members of that maximal set.<sup>30</sup>

Let's see how this works. Suppose we are given the set containing all large sensibles. We know this set is at stage 0, and, further, that it is a maximal set. By a double use of universal elimination on (OM), we can say the following: 'If  $x$  is a maximal set of Fs at stage 0, then there is exactly one form of F at stage  $0 + 1$  participated in by all and only the members of it.' The protasis is true, so there must be exactly one form of F at stage  $0 + 1$  participated in by all and only members of our maximal set of large sensibles at stage  $n$ . Thus, this form of large will be at stage 1. So we now have a maximal set of large sensibles at stage 0, as well as exactly one form of large at stage 1. Consider now the union of the

---

<sup>30</sup> My formulation of (OM) is similar to Fine's. I choose Fine's because I think it works out some of the kinks in Cohen's (OM) – some of which Cohen himself fixed in the paper, but not as concisely as Fine. For instance, Cohen needed to bring in the qualifying adverb 'immediately' to modify the participation relation – which I think Fine more precisely captures by quantifying over the natural numbers, such that she can just say ' $n + 1$ '. But Fine doesn't make explicit the fact that her (OM) is quantified over the natural numbers ( $n$  occurs free): this could be misleading when trying to construct formal proofs involving (OM). She also restricts her (OM) to maximal sets, which, I argue, is problematic.

maximal set of large sensibles with the unique form of large at stage 1. This too is a maximal set, for the following reasons: (i) it contains every F-object at the stage equal to that of its highest-stage member (since the stage of our highest-stage member is 1, and there is *exactly one* [i.e. no more than one] form of F at stage 1), and (ii) it contains every F-object at every stage less than the stage of its highest-stage member (since the only stage less than 1 is 0, and our maximal set contains all F-objects at stage 0). Further, this maximal set is at stage 1, since its highest-stage member is a form at stage 1. By another double use of universal elimination on (OM) (substituting '1' for 'n' this time) and subsequent conditional elimination, we can say that there is exactly one form of large at stage 2 participated in by all and only the members of the maximal set of Fs at stage 1. We can see how this process is already beginning to repeat.

There is at least one problem with the non-terminating derivation sketched above.<sup>31</sup> It is this: (OM) requires that we be given an initial maximal set

---

<sup>31</sup> Cohen himself anticipated this worry (cf. Cohen 463). My solution follows his along roughly the same lines, though my final formulation of (OM) will differ from his. To be sure, there are many other problems with (OM), most of which I won't be able to address in this paper. I'll try to give a brief look into one other problem in this note.

Cohen offers two theorems to solve two of the main problems he finds with (OM); while he gives ample proof, he does not initially tell us precisely why these problems obtain. I'll try to give a set-theoretic explanation of why the first problem obtains (I omit the second worry for the sake of brevity, but at the cost of full clarity, cf. Cohen 463). This first problem turns on a basic principle of set theory: the axiom of extensionality. It tells us that the identity of a set is determined by its members; further, it tells us that repetitions of the members of a set do not affect that set (unlike ordered n-tuples). (We can formalize the axiom of extensionality as such: 'if for every  $\delta$ ,  $\delta \in \alpha$  if and only if  $\delta \in \beta$ , then  $\alpha = \beta$ '. In other words, if two sets have exactly the same members, that is, every member of the first is a member of the second and every member of the second is a member of the first, then they are identical.) So the following expressions designate the same set:

- (a) {George Washington, John Adams}
- (b) {George Washington, George Washington, John Adams}

Obviously, the expression 'George Washington' picks out the same object as 'George Washington.' So the members of (a) and (b) designate the same two objects; thus (a) and (b) are the same set. But how, in our above derivation, did we know that the unique forms of F at stage 1 and stage 2 were *different* forms, i.e. that 'unique form at stage 1' designated a different form than 'unique form at stage 2'? If they were the same, then the union of the maximal set at stage 1 with the unique form at stage 2 would result, simply, in the original maximal set at stage 1, and we could not generate a regress. One might answer by saying that we knew the two forms were different because they were at a different stage. But this just raises the underlying worry: How do we know that objects can't be at more than one stage? One might be initially tempted to bring back (NIF). However, there is an easier way around this worry. We can prove that no object is at more than one stage merely with our working definitions of what it means to be a 'form', to be an 'object', and to be 'at a stage' (cf. Cohen's proof, fn. 28).



---

of Fs in order to generate a new form and thus get the regress going. It is true that we are given some Fs to begin with – this has been an assumption of the TMA all along. But what reason is there for thinking the initial set must be one of Cohen/Fine’s maximal sets? This requirement of (OM) is textually unsatisfying. Plato doesn’t think (OM) – or whatever his explicit one-over-many principle is – needs to obtain only for maximal sets. Rather, his phrasing of (b) simply applies to “*some* number of things” that “seem...to be large” (*Parm.* 132a1; my emphasis). So how do we fix (OM)?

First, we have to remove the term ‘maximal’ from the first part of (OM) – i.e. the protasis within the universally quantified statement. But we don’t want to lose the basic notion behind maximal sets, since that was crucial to our generation of the regress. So we need to build this notion into some aspect of the second half of (OM) – the apodosis. Since the participation relation is included in the apodosis, let’s try formulating it such that we capture the sense of maximal sets. (Note: A subscript  $n$  denotes the stage number of the subscripted variable.)

(Participation–Sets)  $x_{n+1}$  is participated in by  $y_n$  iff  $x_{n+1}$  is participated in by all and only those sets whose stage is less than or equal to that of  $y$  (i.e.  $\leq n$ ).

Note that this is not a definition of participation *as such*. If it were, it would be circular, since ‘participated in’ occurs on both sides of the biconditional; rather, it’s a definition of what it means for a set to participate in something *exactly one stage higher than it*. We can now formulate a better (OM):

(OM<sup>+</sup>) For any  $n \in N$ , and for any set of Fs at stage  $n$ , there is exactly one form of F at stage  $n + 1$  participated in by all and only the members of that set.

(OM<sup>+</sup>) works just as well as (OM) did in validly generating a regress.<sup>32</sup> It does so by retaining the central notion of being a maximal set but relocating it within what I’ll call the ‘ $n+1$  participation’ relation.<sup>33</sup> This relocation allows us to quantify over *all sets* of Fs instead of just maximal sets, thus solving the problem.

#### e. Benefits of (OM<sup>+</sup>)

---

<sup>32</sup> I’ll omit proof (cf. Cohen 467 for a similar proof; though his terminology is different). One might also just run through the same steps I gave for (OM), keeping the definition of (Participation–Sets) in mind.

<sup>33</sup> Cohen seems to do something equivalent, calling his formulation the ‘immediately over’ relation. This is unnecessarily vague. It is better, I think, to spell out precisely what ‘immediately’ means in terms of arithmetic and the natural numbers (hence, my ‘ $n + 1$ ’).

---

We've given a brief and admittedly incomplete sketch of (OM'). To fully justify why (OM') makes the TMA valid, we'd need a more careful presentation of definitions and proofs. But the basic intuition should be evident. So why (OM')? First, it seems to form a consistent triadic premise set with (SPF) and (NIF) (See Item 3 in Appendix).<sup>34</sup> Second, it validly generates a regress given some F-things, as demonstrated in section (d). Third, it stays faithful to the text. This last point is far from obvious.

An objector may be dissatisfied with (OM')'s introduction of set-theoretical and stage-functional concepts on the basis that we do not find any such concepts in Plato's text. However, (OM') seems to be quite faithful to the sense of the text even though Plato, of course, did not frame his TMA in the language of set theory. Plato is clearly discussing *groups* of objects that have some character. Though he may not refer to them as sets, he deploys compound terms like "all these," combined with procedural instructions like, "look at them in the same way," – all of which suggest the process of considering a group of "some" objects as a set of things, taken together. Although Plato doesn't discuss stages, he does elsewhere discuss degrees of reality and thus seems amenable to the notion of hierarchy, especially apropos forms and sensibles, and the participation relation (cf. *Rep.* 515d, *Phd.* 74a). Further, the TMA seems to deploy this notion of orders or stages of objects and forms, albeit not in the spatial sense that 'stage' or 'level' implies; in the text, we find this notion engendered by the "in virtue of" relation.

(OM') is also textually faithful in that it posits 'exactly one' Form over any given set. It thus serves its role, in the argumentative context, to tempt Plato. This is a virtue, as we have seen, that formulations of (b) like that of Sellars lack. Further, (OM') doesn't go too far and build in uniqueness – a virtue that Vlastos's (b) lacked. (OM') *looks* to imply that there is only one form over the many, but upon further thought, it generates a regress of forms.

## II. Vulnerability

Up to this point, all I have done is try to formulate the TMA such that it validly generates a regress from consistent premises and remains faithful to the text. I tried to formulate it as forcefully as possible, so that it poses a formidable

---

<sup>34</sup> Of course, we've shown that we don't really need (SPF) and (NIF). Nevertheless, (OM) should be consistent with them, since they are, as Cohen alleges, built into the definitions of 'form', 'object' and 'at a stage' that form the vocabulary of (OM). (For a fuller discussion, cf. Cohen, 464-5.) Fine keeps them as separate, implicit premises. So either way, I think a justification of the consistency of all three premises (not just the internal consistency of (OM)) is not too much to ask.

threat to Plato's theory. But this does not mean Plato is vulnerable to the TMA. He is vulnerable if, and only if, he is committed to all three premises that we've claimed the TMA (on its best rendering) involves: (SPF), (NIF), and (OM'). But is he? I argue that he is not. To defend the view that Plato is not vulnerable to the TMA, it will suffice to show that he is not committed to just one of the premises. I believe he's not committed to (OM').

A corollary of (OM') is this:

(OMC) Even sets of F-things that include one or more forms of F demand that there be exactly one form (at a stage once greater) 'over' that set.

Without (OMC), the regress would never get going. But we do not find Plato committed to this claim. Consider the "deficiency" argument in the *Phaedo* and the "imperfection" argument in the *Republic* (*Phdo* 74e; *Rep.* 523-5). There, Plato argues that since sensible Fs can appear both F and not-F, they must be deficiently F.<sup>35</sup> This leads him eventually to conclude that there must be a non-deficient – i.e. perfectly F – form of F that is "not the same" as the sensible Fs and in virtue of which the sensible Fs come to be merely deficient Fs (*Phdo.* 74c). This is an argument for the existence of forms: There must be some non-sensible entity – the form of F – that "reports to the soul that the same thing [i.e. some F-object] is perceived by it to be both" F and not-F (*Rep.* 524a). Notice here that it is the *imperfection* of sensible Fs that engenders a (perfect) form of F.<sup>36</sup> We could approximate this formula in set-theoretic, stage-functional terms:<sup>37</sup>

(IMP) For all x, if x is a set of imperfectly F things at stage n, then there is exactly one perfectly F form of F 'over' it (at stage n+1), in virtue of which x is F.

---

<sup>35</sup> This doesn't hold for all predicates substitutable for F – only predicates corresponding to sensible properties like 'large' or 'green' that admit of opposites. (Precisely what 'admit of opposites' mean is debated in the literature. For Fine, Plato considers these properties to suffer from "narrow compresence": "Something suffers narrow compresence with respect to being F if it is F and not F in virtue of some one and the same aspect of itself" (Fine 47, 227)). Predicates like 'man' correspond to sensible properties, but they do not suffer from such compresence (*Rep* 523c-d).

<sup>36</sup> We can phrase this in a number of ways; I'm not sure which is best. Epistemologically: It is the imperfection of sensible Fs that warrants an explanation by means of a perfectly F form of F. Ontologically: It is the imperfection of sensible Fs that posits the existence of a perfectly F form of F.

<sup>37</sup> I try to formulate (IMP) as closely to (OM') and (OMC) as possible, so that anyone who thinks Plato is committed to (OM') will not object that I rigged the Imperfection Argument in my favor. This involves deploying set-theoretic and stage-functional notions central to (OM').

Plato appears committed to (IMP), as *Phdo.* 74-75 and *Rep.* 523-5 suggest. But (IMP) does not imply (OMC). In fact, it implies quite the opposite. Here's why: Take a set of, for example, sensible large elephants. (Note: I will not be discussing their being elephants, only their being large.) This set is at stage 0, by the initial 'iff'-clause of our definition of 'being at a stage'. The members of this set are all imperfectly large. This is because each is both large (at least in some token cases; e.g. when we stand next to it) and not-large (at least in other token cases; e.g. when we see it from a helicopter). By (IMP), there is exactly one perfectly large form of large at stage 1, in virtue of which this set is F. Now consider the set of large sensible elephants in union with this form. This is a set of F things. But it is *not* a set of imperfectly F things, since one of its members is perfectly F.<sup>38</sup> So (IMP), when reapplied to this set, does not generate a new form. It may generate a form of large 'over' the subset of imperfectly large elephants; but this form would be at the same stage as the original form (i.e. stage 1). And since there is exactly one form at each stage, this form would be identical to the one generated by our first application of (IMP). So, from (IMP), we conclude: It is not the case that even sets of F-things that include one or more forms of F demand that there be exactly one form (at a stage once greater) 'over' that set. Which is to say: It is not the case that (OMC). Hence, (IMP), so far from implying (OMC), implies its negation. Now (OMC) is merely an articulation of (OM'), quantifying over *one particular type* of F-set (the type that includes not just sensible Fs, but also forms of F). But (OM') quantifies over *all* F-sets. So if (OMC) is not true, then (OM') is not true. So (IMP) implies that (OM') is not true. I conclude, then, that (IMP) does not just "avoid" (OM') – it expressly denies it (Fine 227).

Our analysis of the Imperfection Arguments in the *Republic* and *Phaedo* seems to demonstrate that Plato is not committed to (OM'). If this is true, then he is not vulnerable to the TMA. However, we have analyzed only one small section of middle dialogues that suggests Plato's noncommittal to (OM'). There is much debate on the issue (cf. Hunt 1997 for a very different take), and many more

---

<sup>38</sup> I am here assuming a definition which we should probably formalize. We have seen what it means for an object to be imperfectly F, but we have not seen what it means for a set to be imperfectly F:

(Imperfection–Sets) a *set* is imperfectly F iff all of its members are imperfectly F. My argument in Part II is much indebted to Fine's ch. 16.2, and it takes much of the same stand. However, I think the validity of her argument for Plato's noncommittal to (OM) – insofar as the Imperfection Arguments are concerned – rests on this underlying assumption, which she does not, to my knowledge, articulate (cf. Fine 227). But is there textual support for our assuming (Imperfection–Sets) holds? Does Plato think this principle is true?

---

passages are worth discussing. In light of this, we'll have to consider the above analysis merely preliminary.

### III. Recursion and the Possibility of Knowledge

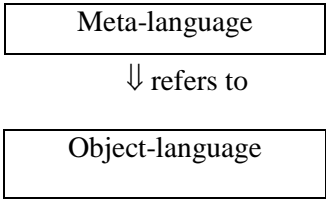
In Part I, we tried to devise a formulation of the TMA such that (a) it contained a consistent premise set, (b) the premises validly entailed a regress, and (c) our premises stayed faithful to the argumentative context in Plato's text. We came up with a reasonable stage-functional formulation of the TMA that appeared, at least *prima facie*, to satisfy these conditions and to pose a formidable threat to Plato. In Part II, we asked the question: Is Plato vulnerable to the TMA as we formulated it in Part I? We reached the very tentative conclusion that he is not. But there is a third topic in need of discussion. Suppose we are wrong, and Plato *is* committed to all the premises of the TMA. Even so, we might wonder: How much of a problem is the TMA? In other words, to what extent does the regress engendered by the TMA actually threaten Plato's theory of forms and, consequently, his claim that knowledge is possible?

I initially envisioned our formulation of the one-over-many principle, as (OM'), to be a double-edged sword. Just as it enabled the most viable, threatening version of the TMA, I thought, so too it might open a way out for Plato. I anticipated that the stage-functional blueprint that (OM') imposed on Platonic ontology would foster a similarly stage-functional, recursive definition of Platonic knowledge. Such a definition would allow for an infinite regress of forms without leading to circularity or viciousness, in the same way recursive theories of truth avoid circularity. After much toil, I admit failure. Let's look, first, at how the TMA is thought to pose a problem for Plato. Second, let's try to pinpoint precisely why I failed at formulating a successful recursive definition of Platonic knowledge.

One reason why the regress is thought to be troubling is that it challenges Plato's claim that knowledge is possible. Plato appears to hold the view that knowledge requires explanation – a *logos*, or account of why things are the way they are (*Meno* 98a). The only way to attain such an explanation is to know the form of what is in need of explanation; all explanations refer to forms (*Phaedo* 96ff.). For instance, I can know that something is large only if I know the form of largeness. (Or, more precisely: If there is a form of *x*, I can know that something is *x* only if I know the form of *x*. This condition accounts for the possibility that Plato may, at times, limit the range of forms to cover some properties and not others.) Since Plato also appears to hold to (some version of) self-predication that implies that any form of *F* is itself *F*, an explanation of the nature of the form

of F must also involve an explanation of why it is F. If, as the TMA has it, the form of F is F in virtue of another form, and that of another, *ad infinitum*, then we would never know how anything is F. So Plato's position, at least initially, seems to be this: Knowledge of F is possible just in case there is some F-thing that is F in virtue of *itself*. But the conclusion of the TMA appears to falsify the latter half of this biconditional and thus challenge the possibility of knowledge in Plato's theory.

Suppose we wanted to vindicate Plato from this vicious regress. We might, taking a cue from the stage-functional apparatus invoked by (OM'), try to formulate a stage-functional definition of knowledge. From this definition, there would arise a manifold of predicates: not just 'KnowledgeOf', but also 'KnowledgeOf<sub>1</sub>', 'KnowledgeOf<sub>2</sub>', ... , 'KnowledgeOf<sub>n</sub>'. At each successive stage of F-objects, there would be a corresponding stage of knowledge. At these higher stages, the Platonist would not *know simpliciter* that  $x_n$  is F, but she could *know<sub>n</sub>* that x is F. This, so the claim goes, would make 'knowledge' (recursively defined in terms of stages) possible. Let's try this tactic, using a familiar recursive theory of truth as our model.<sup>39</sup> On this theory – which I will refer to as Theory T – we distinguish between two levels of discourse.



Semantic theories that adopt T's model tell us that *True* is a predicate in the meta-language, applied to statements in the object-language, such that:

(T) For all x, *True*(x) iff  $\phi(x)$ ,

where there are no occurrences of *True* in  $\phi$ . We arrive at biconditionals such as: "Snow is white" is true iff snow is white. Notice that T is not a circular definition of truth, since 'True' does not occur on the right side of the biconditional. It is recursive, however. You could always have a predicate *Truth<sub>2</sub>*, in a meta-meta-language, which is applied to statements in the meta-language. Put generally, this

---

<sup>39</sup> The following theory of truth is a rough sketch of the one Tarski proposed in his 'The concept of truth in the language of the deductive sciences'. (Reprinted in English, in *Logic, Semantics, Metamathematics, papers from 1923 to 1938*, ed. John Corcoran, Indianapolis : Hackett Publishing Company.).

---

scheme allows for  $\text{True}_n$ , where  $n$  is an  $n$ th-stage meta-language.  $\text{True}_n$  would then be ascribed to statements on the  $(n - 1)$ th-level ‘object language,’ which could itself be a meta-language. So, there may be an infinite regress of *True*’s – an infinite regress of ‘if and only ifs’ – but this regress would not be vicious since *True*, at whatever stage, could be whittled down to the initial stage zero, exemplified by T.

One might think we could construct a roughly analogous model for knowledge. On this model, we would take ‘KnowledgeOf’ to be a predicate in Plato’s meta-language, applied not to statements but to objects on the ground level. So, by rough analogy,

(a) For any object  $x$  and property  $F$ ,  $\text{KnowledgeOf}(x \text{ is } F)$  iff  $x$  is  $F$ .

Now, of course, there may be a predicate ‘ $\text{KnowledgeOf}_2$ ’, in a meta-meta-language one stage higher than ‘KnowledgeOf’, such that

(b) For all  $x$  and property  $F$ ,  $\text{KnowledgeOf}_2(x \text{ is } F)$  iff

$\text{KnowledgeOf}(x \text{ is } F)$

But while this structure makes referential, semantic theories of truth like (T) possible, it doesn’t seem to work for a Platonic theory of knowledge. Any recursive definition relies on at least one ‘base clause,’ which states outright that certain cases instantiate the property or relation being defined. And we are supposed to take this base clause at face value. The ‘induction clause’ or clauses then tell us that, given certain things instantiating the property or relation in question, certain other things also instantiate this property or relation. The induction clauses ‘build up,’ so to speak, from the base clause, forming infinitely many sets at successively higher stages. But, while the successive stages of  $F$ -objects ‘build up,’ on Cohen’s definitions, from the initial

(1) Let an object be at stage 0 iff it is a sensible object,

the successive stages of the ‘KnowledgeOf’ relation *do not* ‘build up’ from any solid foundation. Rather, the ‘base’ clause that grounds Plato’s ‘KnowledgeOf’ relation just introduces *another* relation – namely, the ‘ParticipatesIn’ relation – which holds between objects at stage 0 and objects at stage 1. It turns out, then, that our base clause for the ‘KnowledgeOf’ relation (i.e. what should be an outright statement of what counts as an instance of ‘knowledge’ at stage 0) *already includes reference to objects at stage 1*. But this does not fit the bill of a proper base clause for a recursive definition, and we can’t just take it at face value. We end up not ‘building up’ from this base, but ‘descending down,’ in an

---

effort to explain the elusive ‘ParticipatesIn’ relation which, we are told, holds between objects *at two different stages*.<sup>40</sup>

---

<sup>40</sup> A spatial analogy: Each stage of a recursive definition is like a transparent floor in a building. Suppose we are on the fifth floor. The ceiling on the fifth floor is made of mirror, such that when we look up, we see a reflection of ourselves. But we also see a reflection of every floor below us. This is because the tile we are standing on is transparent, and the tile that fourth-floor people are standing on is transparent as well, and such is the case all the way down to the ground floor. So, in effect, there is a two-way mirror separating each floor of this building; you can see one way (down), but you can’t see the other way (up). When fifth-floor people like us look down, we see that we are just like those on the fourth floor, except there’s one new thing about us (say, we all have red hats). We are, in this sense, impure compared to those without this addition. And the fourth-floor people, bereft of red hats (indeed, ignorant of even the existence of red hats), look down and realize they are just like the third floor people, except they too have some one new thing about them (striped shirts), and are, in this way, impure as well. And so on for the third- and second-floor people. But the ground floor people look down and just see brick. They look up and are met with a reflection of themselves. They think themselves all alone. They don’t know why they are the way they are; nothing has been added to them. They are just plain people.

But Plato’s Knowledge Building has a completely different blueprint from the People Building. Everything is inverted. The two-way mirrors that separated floors have been replaced with glass. Thus, knowledge-0 on the ground floor can look up and see that it is, not the pure version (no additives) of knowledge at the second floor, but the *impure* version of Knowledge-2 on the second floor – and so on up the building. The main point is this: Plato’s base clause – his knowledge-0 at the ground level – is not cut off from knowledge-2 at the second level: Rather, it is intertwined, by definition, with the knowledge of the higher levels. Ground level knowledge is already dependent on level-two knowledge.



---

It is this suture between objects at two different stages that throws off the recursive model. This is why I was unable to formulate a recursive theory of Platonic knowledge based on the stage-functional apparatuses at work in (OM<sup>\*</sup>). The difficulty centers on the absence of a proper base clause and the *externality* of the ‘ParticipatesIn’ relation to any one stage in particular.

#### IV. Concluding Remarks

In Part I, we tried to formulate the most forceful and textually appropriate version of the TMA’s premises. As such, the TMA looked to be a formidable opponent for Plato. In Part II, we found that Plato did not seem committed to one of the premises – (OM<sup>\*</sup>) – that we formulated in Part I. But this proof of Plato’s lack of commitment was incomplete at best, since it only addressed a select few passages from the *Republic* and *Phaedo*. In Part III, we assumed the worst: Suppose our incomplete proof in Part II is wrong, and Plato *is* committed to the premises sketched in Part I – i.e. suppose Plato is vulnerable to the conclusion of the TMA. We asked: Is the conclusion of the TMA actually problematic? We tried to answer in the negative by formulating a recursive definition of Platonic knowledge, by analogy to the stage-functional scheme of (OM<sup>\*</sup>), but hit what seemed to be an insurmountable wall. The multiplicity of our conclusions here seems to testify to the peculiar nuance of Plato’s Third Man Argument and the manifold issues at stake in that small passage in the *Parmenides*.

#### Appendix

##### I. Item 1

- a. Claim: (SPF) and (NIF) are consistent.

Note: This claim is almost self-evident, and it is mostly accepted in the literature (Cohen 455, Fine 207). However, since I do not defend it in the paper (and since we do not find it formally defended in the literature), I will give a model-theoretic justification here. If (SPF) and (NIF) are consistent, then we should be able to give a model making true {(SPF), (NIF)}. First, we’ll need some notation.

---

Let 'FF' = a one-place predicate standing in for 'is a form of F'

Let 'F' = a one-place predicate standing in for 'is F'

Let 'P' = a two-place relation standing in for 'participates in,' or 'is F in virtue of'

Let  $U$  = the universe of the model

Let  $V$  = the valuation function of the model

Let  $S$  = the signature set of the model

Let the gothic 'a' denote the model

With this notation, I translate (SPF) and (NIF) into first-order formal language as such:

(SPF)  $\forall x(FF(x) \supset F(x))$

i.e. 'for every x, if x is a form of F, then x is F'

(NIF)  $\forall x(F(x) \supset \forall y((FF(y) \wedge P(x,y)) \supset x \neq y)))$

i.e. 'for every x, if x is F, then for every y, if y is a form of F and x is F in virtue of y, then x is not identical with y'

Here's some notation external to our formal language that will save some ink:

Let left facing arrows denote 'because'

Let double turnstiles denote 'makes true'

[See subsequent page for model]

$$S = \{ 'FF^2', 'F^2', 'P^2' \}$$

$$\sigma \left\{ \begin{array}{l} U = \{ 2, 3 \} \\ \mathcal{V}('FF^2') = \{ 3 \} \subseteq U^2 \\ \mathcal{V}('F^2') = \{ 2 \} \subseteq U^1 \\ \mathcal{V}('P^2') = \{ \langle 3, 2 \rangle \} \subseteq U^2 \end{array} \right.$$

$$(SPF) \quad \sigma \models \forall x (FF(x) \supset F(x))' \leftarrow \textcircled{1} \sigma \models 'FF(x) \supset F(x)' \leftarrow \sigma \not\models 'FF(x)'$$

$$\leftarrow 'x'^{\sigma} = 2 \notin \mathcal{V}('FF') \leftarrow 2 \notin \{ 3 \}$$

$$\textcircled{2} \sigma \models 'FF(x) \supset F(x)' \leftarrow \sigma \models 'F x'$$

$$\leftarrow 'x'^{\sigma} = 3 \in \mathcal{V}('F') \leftarrow 3 \in \{ 3 \}$$

$$(NIF) \quad \sigma \models \forall x (F(x) \supset \forall y ((FF(y) \wedge P(x, y)) \supset x \neq y))' \leftarrow$$

$$\textcircled{1} \sigma \models 'F(x) \supset \forall y ((FF(y) \wedge P(x, y)) \supset x \neq y)' \leftarrow$$

$$\sigma \models '\forall y ((FF(y) \wedge P(x, y)) \supset x \neq y)' \leftarrow$$

$$\textcircled{i} \sigma \models '(FF(y) \wedge P(x, y)) \supset x \neq y' \leftarrow$$

$$\sigma \not\models 'FF(y) \wedge P(x, y)' \leftarrow$$

$$\sigma \not\models 'FF(y)' \leftarrow 'y'^{\sigma} = 2 \notin \mathcal{V}('FF')$$

$$\textcircled{ii} \sigma \models '(FF(y) \wedge P(x, y)) \supset x \neq y' \leftarrow$$

$$\sigma \models 'x \neq y' \leftarrow 'x'^{\sigma} \neq 'y'^{\sigma} \leftarrow 2 \neq 3$$

$$\textcircled{2} \sigma \models 'F(x) \supset \forall y ((FF(y) \wedge P(x, y)) \supset x \neq y)' \leftarrow$$

$$\leftarrow \sigma \not\models 'F(x)' \leftarrow 'x'^{\sigma} = 3 \notin \mathcal{V}('F')$$

ITEM 2:

II. Item 2

a. Claim: (bV) combines with (SPF) and (NIF) to form an inconsistent triad.

Note: The force of (bV) can be summed up in quantificational terms, as Cohen notes, in the following way: 'for any x, if x is F then there is exactly

---

one F-ness in which x participates' (Cohen, 458). I give a formal derivation showing that the absurd is a logical consequence of  $\{(bV), (SPF), (NIF), (F(a) - \text{i.e. some F-thing})\}$ . I do so in the Fitch format. My purpose is threefold: (1) to substantiate a claim I make in the paper; (2) to demonstrate that the derivation of a contradiction from these three premises is not as straightforward as it might initially appear, or, as informal proofs make it seem (c.f. Fine, 208) – it takes two applications of Vlastos's (b) before we arrive at contradictory claims (i.e. a double use of existential elimination); and (3) to show, conclusively, that we cannot produce a satisfactory formulation of (b) so long as it quantifies over Fs and not *sets* of Fs.

As in Item 1, I translate (SPF) and (NIF) into first-order formal language (no changes). Further, I translate (bV).

(bV)     $\forall x(F(x) \supset \exists y(FF(y) \wedge P(x,y) \wedge \forall z(FF(z) \supset y = z)))$   
           i.e. 'for any x, if x is F then there is exactly one form of  
           F in which x participates'

[See subsequent page for derivation]

1	$\forall x (FF(x) \supset F(x))$	
2	$\forall x (F(x) \supset \forall y ((FF(y) \wedge P(x, y)) \supset x \neq y))$	
3	$\forall x (F(x) \supset \exists y (FF(y) \wedge P(x, y) \wedge \forall z (FF(z) \supset z = y)))$	
4	$F(a)$	
5	$F(a) \supset \exists y (FF(y) \wedge P(a, y) \wedge \forall z (FF(z) \supset z = y))$	$\forall E: 3$
6	$\exists y (FF(y) \wedge P(a, y) \wedge \forall z (FF(z) \supset z = y))$	$\supset E: 4, 5$
7	$FF(u) \wedge P(a, u) \wedge \forall z (FF(z) \supset z = u)$	
8	$\forall x (FF(x) \supset F(x))$	$R: 1$
9	$FF(u) \supset F(u)$	$\forall E: 8$
10	$FF(u)$	$\wedge E: 7$
11	$F(u)$	$\supset E: 9, 10$
12	$\forall x (F(x) \supset \exists y (FF(y) \wedge P(x, y) \wedge \forall z (FF(z) \supset z = y)))$	$R: 3$
13	$F(u) \supset \exists y (FF(y) \wedge P(u, y) \wedge \forall z (FF(z) \supset z = w))$	$\forall E: 12$
14	$\exists y (FF(y) \wedge P(u, y) \wedge \forall z (FF(z) \supset z = w))$	$\supset E: 13, 11$
15	$FF(w) \wedge P(u, w) \wedge \forall z (FF(z) \supset z = w)$	
16	$\forall x (FF(x) \supset F(x))$	$R: 1$
17	$FF(w) \supset F(w)$	$\forall E: 16$
18	$F(w)$	$\supset E: 15, 17$
19	$F(u)$	$R: 11$
20	$F(u) \supset \forall y ((FF(y) \wedge P(u, y)) \supset u \neq y)$	$R: 3, \forall I: 3$
21	$\forall y ((FF(y) \wedge P(u, y)) \supset u \neq y)$	$\supset E: 19, 20$
22	$(FF(w) \wedge P(u, w)) \supset u \neq w$	$\forall E: 21, 20$
23	$FF(w) \wedge P(u, w)$	$\wedge I: 15$
24	$u \neq w$	$\supset E: 22, 23$
25	$\forall z (FF(z) \supset z = w)$	$\forall E: 15$
26	$FF(u) \supset u = w$	$\forall E: 25$
27	$FF(u)$	$R: 10$
28	$u = w$	$\supset E: 26, 27$
29	$\perp$	$\supset E: 24, 28$
30	$\perp$	$\exists E: 14, 15-29$
31	$\perp$	$\exists E: 6, 7-30$
32	[enter what you like: conclusion of TMA]	$\perp E: 31$

### III. Item 3

a. Claim:  $(OM')$  forms a consistent triadic premise set with  $(SPF)$  and  $(NIF)$ .

Note: We cannot proceed in justifying the claim that  $(OM')$  is consistent in the same way as we did for Item 1. This is because  $(OM')$  does not fall under the scope of the first-order logic and model theory we've been using up to this point. Instead, we need to think in terms of variables ranging over, not

---

objects in the universe, but *predicates* in our vocabulary set. It quickly becomes evident that there is a problem with (OM<sup>\*</sup>): it is inconsistent, as it stands, so long as we imagine the possibility of a set consisting of every F-thing (either form or sensible). This possibility is captured in the principle of abstraction, which says that for every predicate, there is some set,  $\alpha$ , containing all and only objects to which that predicate applies. Suppose there is such a set,  $\alpha$ , containing every F-thing. On our stage-functional scheme, this set would have no highest-stage member, since it would be an infinite set containing members of infinitely high(er) stages. Hence, we could not say this set is at any stage (on our definition of what it means to be ‘at a stage’). Now subbing this infinite set into (OM<sup>\*</sup>), we get the claim that there is exactly one form of F at stage  $n + 1$  participated in by all and only the members of that set. Suppose there is such a unique form ‘over’ this set. But this unique form, then, must not be a member of  $\alpha$ , since it is ‘over’ it. But the form is itself F (by (SPF)); and  $\alpha$  contains every F-thing. So this form is both a member of  $\alpha$  and not a member of  $\alpha$ . Hence, inconsistency.

Both Fine and Cohen answer this problem posed by the principle of abstraction. Fine assumes from the start that “every maximal set is at a level” (Fine n. 42). Cohen builds this thought into his second, more precise formulation of (OM). He does so by quantifying not universally over the natural numbers, but existentially: ‘there is some  $n$ ’ (Cohen n. 33). I wonder why we they feel compelled to make these alterations. It seems like a set that includes an infinite amount of members (at infinitely high stages) would not be applicable to (OM<sup>\*</sup>) in the first place. Recall that the first quantifier in my formulation of (OM<sup>\*</sup>) is a universal, over the natural numbers. What would we substitute in for this  $n$ , in order to make (OM<sup>\*</sup>) susceptible to such an infinite set – a set which doesn’t even have a finite stage number? Infinity? Is infinity a *member* of the set of natural numbers? Perhaps the set of natural numbers is infinite, but that does not necessarily mean infinity *itself* is a member therein. Further, is infinity really the *stage number* of this infinite set of Fs? Not on our definition of what it means to be ‘at a stage’. There simply is no stage number; no  $n$  to substitute into (OM<sup>\*</sup>). Does this imply a contradiction? I don’t think so. We only arrived at a contradiction *once* we eliminated the universally quantified  $n$ , and, subsequently, got a fresh form one stage higher than our big peculiar set. But if we can’t even universally eliminate, we can’t get to the conditional. So, while I see the problem the principle of abstraction poses, I don’t see precisely how this problem results in an inconsistency.

---

Even if we avoid the inconsistency charge posed by the principle of abstraction (either by Cohen's or Fine's way, or by my pseudo-suggestion), we have not sufficiently defended the view that (OM'), (SPF) and (NIF) are consistent; all this has shown is that our premises survive *one* attempt to derive a contradiction from them. Certainly, however, (OM') is not, *prima facie*, inconsistent (as Vlastos' (b) was). So this is a virtue. Nevertheless, I conclude that this topic deserves much more attention than it seems to have received in the literature.

---

## Works Cited

- Cohen, S. Marc. "The Logic of the Third Man," *The Philosophical Review*, Vol. 80, No. 4 (Oct., 1971). Pp. 448-475.
- Fine, Gail. *On Ideas: Aristotle's Criticism of Plato's Theory of Forms*; Clarendon Press, Oxford: 1993.
- Hunt, David P. "How (Not) to Exempt Platonic Forms from "Parmenides"" Third Man," *Phronesis*, Vol. 42, No. 1 (1997), pp. 1-20.
- Plato, *Parmenides*, trans. Mary Louis Gill and Paul Ryan. In Plato: Complete Works, ed. Cooper, 1997; Hackett Publishing.
- Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo, trans. G.M.A. Grube. Hackett Publishing, 2002.
- Republic*, trans. G.M.A. Grube, Hackett Publishing 1992.
- Sellars, Wilfrid. "Vlastos and 'The Third Man,'" *The Philosophical Review*, Vol. 64, No. 3 (Jul., 1955), pp. 405-437
- Tarski, Alfred. 'The concept of truth in the language of the deductive sciences'. [Reprinted in English, in *Logic, Semantics, Metamathematics, papers from 1923 to 1938*, ed. John Corcoran, Indianapolis : Hackett Publishing Company].
- Vlastos, Gregory. "The Third Man Argument in the Parmenides," *The Philosophical Review*, Vol. 63, No. 3 (Jul., 1954). Pp. 319-349. [Referred to as Vlastos' first paper, or Vlastos I]
- "Plato's 'Third Man' Argument (PARM. 132A1-B2): Text and Logic," *The Philosophical Quarterly*, Vol. 19, No. 77 (Oct. 1969), pp. 289-301. [Referred to as Vlastos' second paper, or Vlastos II]
- (In addition to the works cited in this paper, I have benefitted from discussion with Gail Fine and Harold Hodes.)



---

## Interview with Daniel Dennett, Tufts University

Daniel Dennett is the Austin B. Fletcher Professor of Philosophy and Co-director of the Center for Cognitive Studies at Tufts University. His research has centered on philosophy of science and philosophy of mind, with particular interests in cognitive science and evolutionary biology. He is a steadfast and vocal atheist and secularist. His many books have been read widely both in and out of the academy. They include *Consciousness Explained* (1992), *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (1996), and *Breaking the Spell: Religion as a Natural Phenomenon* (2006).

We spoke with Professor Dennett in June 2010.

**YPR:** *Can you explain the distinction you've made between content and consciousness, and why you think it's a useful distinction?*

**DANIEL DENNETT:** I think I have always been interested, from the time I was an undergraduate at Harvard if not before, in questions about the “sources of normativity,” as Christine Korsgaard calls it: What can make it the case that we ought to do something and, in particular, that we are morally obligated to act? My dissertation as a graduate student at the University of Pittsburgh was about (normative) reasons for acting: What makes some consideration a reason for somebody to do something? And I was always interested in issues about the nature and authority of morality: What kind of reasons for acting do moral obligations purport to provide and do these reasons really exist? I was very lucky in graduate school and my early career to be exposed to a number of philosophers who had thought hard about these issues—Kurt Baier, William Frankena, and W. D. (David) Falk, who was my colleague at the University of North Carolina. My first book, *Impartial Reason*, brings together whatever progress I had been able to make on these questions in the decade after I finished graduate school. One thing I am pretty proud of is that whereas the orthodox view, which I was 1. Content is meaning—for instance, intentional content, propositional content, conceptual content. In traditional views, it's what comes after the “that” in a ‘that’ clause, and it has nothing to do with consciousness. There are lots of systems that have content without consciousness. I realized that back at the beginning of my career, when I was a graduate student: the trick was to do a

---

theory that could handle content first. Then you'd see that consciousness was a later sophistication—an add-on—that elevates some of the content to a special functional role. That special role is consciousness.

To take an example, say that you're driving down the road and there are some tree branches waving in the wind. Compare that to when you're driving down the road and a person is waving. The distinction between those two situations—the difference in the content of your mental state—is precisely what makes you conscious (probably) of the latter [the person waving] rather than the former. The content determination, the content fixation, is done by relatively early processes in your nervous system, and on the basis of the content that is at first crudely, and then more subtly, determined, things advance to consciousness or not. We're conscious of only a tiny fraction of the contents that are being distinguished by our nervous systems.

**YPR:** *As a philosopher of mind, what level of attention do you pay to advances in cognitive science and neurobiology? Is there an example of an empirical finding that directly inspired one of your philosophical ideas?*

**DD:** I pay as much attention as I can: I think that's the only way to do philosophy of mind. I think that the attempt to work out interesting analyses and claims from first principles is almost always forlorn—just a waste of time and energy. At a time when cognitive neuroscience, and cognitive science more generally, are not just making interesting discoveries but advancing philosophically very interesting and often very dubious claims, it's a great time for philosophers of mind to pay attention. They can actually help sort out the issues, they can help clarify the issues, because cognitive neuroscientists make a lot of philosophical mistakes, just as everyone does. So that's the way to do philosophy of mind. The philosophers who are unabashedly ignorant of neuroscience, ignorant of cognitive science, but insist that they're philosophers of mind: I think they should be embarrassed; they're working blindfolded.

In 1992, I co-authored a paper with Marcel Kinsbourne called "Time and the Observer: The Where and When of Consciousness in the Brain." That paper formed the backbone of *Consciousness Explained* [Dennett's 1991 book]. And it grew quite directly out of my considering [neurophysiologist Benjamin] Libet's results and thinking that they had been completely misinterpreted by Libet and many others, and trying to think about what a good interpretation would be.

In his so-called free will experiments, Libet had people spontaneously flick their wrist whenever the spirit moved them, for no reason at all, while

---

looking at a rapidly rotating disk with a dot on it, so that they could say, ‘Well, I decided to flick my wrist when the dot was at 3 o’clock, or 4 o’clock, or 5 o’clock.’ So the subject is timing his subjective consciousness of when he decided to flick his wrist. And meanwhile, Libet was measuring the *Bereitschaftspotential*, or readiness potential, which is a characteristic brain wave that had been detected by [the neuroscientists] Kornhüber and Deecke about 10 years earlier. Libet found that the potential was occurring two or three hundred milliseconds ahead of the time when subjects said they thought they were conscious of deciding. In other words, the brain was deciding two or three hundred milliseconds before the mind was conscious of deciding, which suggested that free will was an illusion, because the consciousness of decision is lagging by about a quarter of a second behind the actual decision-making that’s going on in the brain.

That’s the standard interpretation of Libet. And it’s a mistake. There are many things wrong with that interpretation, but I’m not going to try to say just exactly what they are, because I’ve written at great length about this. But just to give you a small hint of what’s wrong with the interpretation, suppose that political scientists put out various probes and predict that a particular bill will pass the Senate in a couple of weeks. And suppose the predictions they make on the basis of the evidence they gather with their probes turn out to be highly accurate. Would that show that the Senate had already voted and already decided? No, it wouldn’t show that at all. Even if you can predict what they’re going to decide, it doesn’t follow that they have already decided it. That’s just a tiny part of what’s bizarre about the standard interpretation of Libet, which is very Cartesian. The good thing about the Libet experiment is that it’s a great closet-emptier: It brings the Cartesians out of the closet just beautifully. People hear the results and they say, “Ooh, my gosh, I guess I don’t have free will.” That’s because they’re thinking that free will works this way: You can’t start thinking about making a free decision until you’re conscious of the issue, and *then* you make the decision consciously. That decision must happen in some special privileged chamber of the brain, and if that chamber is in effect out of the loop, then free will is an illusion. This idea that there’s a place in the brain where consciousness happens is just a big mistake. That’s Cartesian materialism.

**YPR:** *In your experience, what are some of the roles taken on by philosophers of mind that empirical researchers either cannot or choose not to take on?*

---

**DD:** I think there are several. [Harvard Professor] Peter Godfrey Smith, in his most recent book about biology and evolution, *Darwinian Populations and Natural Selection*, nicely distinguishes three tasks for philosophers of science. One is simply to do philosophy of science for philosophers of science—that is, to take on the internal disputes and disagreements. It’s not terribly important for scientists to read that work. The second is to actually attempt to make a contribution to the science by clarifying concepts, showing that certain arguments that have been run are inconclusive or are conclusive. That’s analogous in a way to what theoretical physicists do. And the third goal that Godfrey Smith describes is what he calls philosophy of nature, which is working out the larger implications of the science: in other words, explaining to the general public what the import is of various discoveries and advances and suggestions from science. I think that’s a pretty good trio, and I agree with Godfrey Smith that philosophers should aspire to all three. The only way they can do that is if they are really quite deeply knowledgeable and acquainted with the science in question—and the philosophy in question.

I think, for instance, when scientists decide to take on the public role and do the philosophy of nature, they often make a hash of it, in particular on the free will issue. There have been a number of really distinguished neuroscientists and psychologists in the last few years who have been going around saying, “Neuroscience shows that we don’t have free will.” Well, they have very crude ideas about what it would be to have free will, and what they’re saying is actually causing quite a stir. There have been some deeply misleading claims that have been made by neuroscientists who think there’s nothing tricky about the issue of free will—but of course it is a tricky issue. So I think philosophers have a major role to play in sorting that out. Free will is an issue that really matters to people and should matter to people.

**YPR:** *You’ve asserted that qualia don’t exist in any meaningful sense of the word and that people cling to them simply in order to retain something ‘special’ that is outside of the reach of science. [‘Qualia’ is the term that some philosophers use for the purported intrinsic, personal properties of consciousness, or as Professor Dennett put it, the “ways things seem to us.”] Many people disagree with you, but it seems that the arguments on either side often take the existence or non-existence of qualia as a premise, so that the existence claim never gets a full hearing. How do you see the debate over qualia going forward in a meaningful way?*

---

**DD:** Heaven knows I've tried to change the terms of the debate, trying to get people off the standard account of qualia and to recognize that 'qualia' is just a poor term, because people don't agree on what it means. I've tried to propose some alternative ways of talking about subjective experience and its properties. Of course, I'm not saying that there aren't itches and tingles of beautiful subjective shades of blue or any of that: of course there are. But there aren't the sorts of magical, trumped-up, inflated properties that philosophers often characterize those subjective properties as being. I think some philosophers take it as definitional that qualia have to be whatever it is that functionalism [the view that an agent's mental states are nothing more than the roles they play for the agent] can't handle. Well, if you define qualia that way then of course you defeat functionalism: you defeat it by fiat. This temptation to assume that no functional decomposition of qualia is possible has profoundly negative effects on thinking in the field of philosophy of mind.

I'm reminded here of my favorite quote from [Professor of Religion] Lee Siegel's book on magic [*Net of Magic: Wonders and Deceptions in India*].

“‘I’m writing a book on magic,’ I explain, and I’m asked, ‘Real magic?’ By *real magic* people mean miracles, thaumaturgical acts, and supernatural powers. ‘No,’ I answer: ‘Conjuring tricks, not real magic.’

“*Real magic*, in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is *not real magic*.” (p.425)

Well, for a lot of people, consciousness has to be real magic. In any theory where it turns out that consciousness is just a bag of tricks—which is what I claim—some people say, “Oh, well you’re just saying that consciousness doesn’t exist.” Well, in a way, yeah. If by definition consciousness is real magic, I’m saying it doesn’t exist. That doesn’t mean that we don’t have subjective experience, that we don’t suffer, that we don’t have pangs of joy and lust and all the rest. It just means that the interpretation of those experiences doesn’t have the metaphysical implications that a lot of philosophers insist that it does.

**YPR:** In your book *Freedom Evolves*, you hint at the position that all of philosophy is an attempt at “memetic engineering,” that is, spreading and gaining support for cultural features, ethical theories, systems of government, and other “memes” that are competing on an evolutionary stage. In your

---

*opinion, what are some of the traits that have made ideas particularly fit for survival and reproduction during the time that you've been working and writing?*

**DD:** We can take some of the relatively superficial features from evolutionary biology itself, and we can say that ease of replication is really important. That's why it's good to have mnemonic crutches built into the vehicles of your ideas. You want to have a memorable, apt title for your theory, a memorable name for your principle. You want to make your idea as clear and unforgettable as possible, as vivid as possible. So that's the superficial side of it. But it's important to remember that side, because no matter how true or beneficial a proposition or an idea is, if it is utterly forgettable and confusable, then it will not manage to play the role that you want it to play. Truth is of course a very nice property for an idea to have—especially truth that is important and relevant to the people who are the vectors of the idea. The folk theories about how true and useful ideas will beat out falsehood and useless ideas—there's a lot of truth to those theories, at least in most cases. But what's interesting are the exceptions: Under what conditions can palpable falsehoods or seriously deleterious ideas spread (the presumption being that they shouldn't)? Some of this is fairly obvious, too: Ideas that are “too good to be true” are often deeply appealing to our anxieties or felt needs, and hence persist in the face of rebuttal. Or just “intriguing” ideas—urban myths, shocking rumors, conspiracy theories. I think people are becoming much more sophisticated and devious memetic engineers—spin doctors, propagandists, viral advertisers—and we have to foster equally sophisticated counter-measures, protecting the minds of the citizenry from these assaults. It's an evolutionary arms race.

**YPR:** *You're well-known as a strong supporter of naturalism, the doctrine that there is no supernatural force acting on the universe. Do you think that science can provide sufficient evidence for you to rationally hold that position, or does naturalism require—at the least given the current stage of science—a leap of faith, perhaps along the lines of trusting in Occam's razor principle [the principle that the simplest explanation for a given phenomenon is the mostly likely to be true]?*

**DD:** It's a bit like Occam's razor—it's an issue of burden of proof. For instance, I don't think it takes much of a leap of faith to believe that there are not any ogres living in the moon. I can't prove that there aren't, in the same way that I can't prove that there's no God. But there is no question, no well-formed scientific

---

question, to which God's existence is a good explanatory answer—or at least I certainly haven't encountered it. You would think that somebody would have come up with a good role for God to play in dissolving a mystery or closing a chasm, but there is no such role. And of course one of the problems is that most definitions of God are in a certain sense self-defeating for that role. They define God either so nebulously or so powerfully that God becomes the perfect wildcard. Something that can explain everything doesn't explain anything.

**YPR:** *What do you regard as the best justification for being religious in the 21<sup>st</sup> century?*

**DD:** I think there's no question that the best, although imperfect, justification for being religious is that it fends off despair—but only, of course, if you can manage to maintain the somewhat self-delusional state of being religious. Now, I don't think that despair is a rational response to the non-existence of God. I just think that for many people, that's the conclusion they jump to [i.e., that despair is the rational response]. And so, it is very much in their interest to cling to the straw, that is, attempt to maintain a belief in God in order to ward off nihilism. But I think the whole issue is mischaracterized by the assumption that nihilism is the only alternative. For those who make that mistake, one can see why they are so desperate to cling to a belief that otherwise has nothing going for it.

Personally, nihilism doesn't present itself as a problem for me. I find the fact that I am fortunate enough to be alive on this amazing planet, in the middle of this amazing universe, to be so wonderful. I see around me so many things to be grateful for, and so much work to do to make the planet in some small way better for others. There's no end of meaning when I consider all the benefits I've received from the efforts of others. There's no end of meaningful work to be done. And I think the fact that we don't have a life after death is a pretty good reason for making the most of the life we have.

**YPR:** *How have your philosophical views changed over time?*

**DD:** I've certainly changed my views on homuncular functionalism and homunculi [the view that the mind is composed of sub-personal agents—each with its own mental repertoire—which cooperatively perform all mental functions]. My change of view is almost comical. If you look in *Content and Consciousness* [published in 1969], you see that I say at one point, "Replacing the little man in the brain with a committee doesn't seem like progress." Oh, yes

---

it is—it is indeed progress. I acknowledged that error many years ago, when I started talking about the regress of homunculi [the idea that the homunculi can be conceptually arranged in a hierarchy, with those on the lower levels having simpler roles].

But then I made another mistake. I underestimated the literalness that one could employ in speaking about the subparts of human beings as little agents. In a passage that's often quoted, I talk about how cognitive science proceeds by taking a whole person and then dividing the person up into a handful of smaller interacting sub-personal agents. Each of those agents is an intentional system, and if you look inside them, it's sort of like nested Russian dolls: you find still smaller groups of agents. And by the time you get down to a neuron, you've got something that could be replaced by a machine. This is how you have a finite regress of homunculi. Well, I still believe in the finite regress, but I think that the individual neurons are much more agent-like—much more homuncular—than I was ever thinking of them before. They are in fact quite clever and selfish little fellows. And understanding how their own agendas—the agendas of individual neurons—play a role in cognition, and particularly in the emotional control of cognition: that is, I think, the key to making progress at this point. It gives you a completely different idea about the computational architecture of the brain, for instance.

**YPR:** *We understand that you've been sailing since you were a child. Can you describe one of the most eventful moments on a sailing trip that you've taken?*

**DD:** Well, there are so many moments, but I'll describe a recent one: Sailing through the night between Mount Desert Rock and Monhegan Island [both off the coast of Maine], and having dolphins accompanying us—me and some of my former graduate students—in the moonlit night. The dolphins surfaced and even leapt along the sides of the boat. And there was a lot of phosphorescence, a lot of bioluminescence, so these dolphins were like comets in the water. It was very beautiful.

**YPR:** *Do you find yourself thinking a lot about philosophy while you're on the water, or do you use sailing as a mental break?*

**DD:** I view sailing in general not as an antidote to philosophy so much as a corrective, of the following sort: We philosophers make lots of mistakes, but our mistakes have very few consequences, aside from a little embarrassment. One



---

can see this at a glance by considering philosophers' lack of need for malpractice insurance. What sort of calamity could befall us as a consequence of one of our philosophical mistakes? But when you're out on a sailboat, your knowledge is really put to the test. So I suppose this is my pragmatism coming out: I really appreciate the fact that I can be out in my boat with friends, and conditions are challenging, and because I know what I'm doing, everybody's safe. If I didn't know what I was doing, it would be really, really dangerous—there are a lot of big, dangerous forces out there. I guess I enjoy the gratification of using my knowledge to the practical end of saving my life and the lives of my crew in relatively risky situations.

But also, I've found that sailboats are great places for really extended philosophical discussions. The last few years I've been organizing what we call "Cognitive Cruises," where my postdocs and I take a cognitive scientist as our guest out on the boat. First, we spend a day or two on shore reading papers and formally introducing a bunch of topics and questions and issues. And after we've had our seminars, then we just get on a boat and go sailing for a few days, and the conversation just happens. There's no schedule, there's no "Now it's somebody's turn to say something": We just go out and sail, and as the conversation develops, sometimes we talk about the theory, and sometimes about the weather or the sailing or the fish, or whatever. But a lot of the time the talk is not just good philosophical talk, but very constructive as well. So this has become a fixture of my summer.

---

#### SPECIAL THANKS

The Yale Philosophy Department

This issue of THE YALE PHILOSOPHY REVIEW was printed by TYCO Copying and Printing of New Haven, Connecticut. The typeface is Times New Roman.

All contents © 2010 THE YALE PHILOSOPHY REVIEW. Reproduction either in whole or part without written permission of the Editors-in-Chief is prohibited. While this journal is published and staffed by students of Yale College, Yale University is not responsible for any of its contents.

