Aporia is published twice each year by the Department of Philosophy at Brigham Young University and is dedicated to recognizing exemplary philosophical work at the undergraduate level. The spring issue each year includes the winning essays of the annual David H. Yarn Philosophical Essay Competition, which is open to BYU undergraduates only, along with other outstanding philosophical essays submitted by undergraduate philosophy students from around the world. The winners of the Yarn Competition are selected by a faculty committee independent of the student editors and staff, who make all other publishing decisions. The fall issue is published online and consists of papers selected by the editorial staff. Submissions are welcome from all undergraduate students, both at BYU and elsewhere.

The Aporia staff is especially grateful to those who support makes this journal possible. Special thanks go to Professor Daniel Graham, chair of the Department of Philosophy, Professor David Jensen, our faculty advisor, and to Linda Hunter Adams of the Humanities Publication Center. Aporia is funded by a generous contribution from the David H. Yarn Endowment. Those wishing to contribute to this fund may do so by contacting the BYU Department of Philosophy.

Opinions expressed herein are those of the respective authors, and do not necessarily reflect the opinions of the editors, the Philosophy Department, or Brigham Young University.

http://humanities.byu.edu/philosophy/aporia/home.html
In his article “On Referring” P. F. Strawson argues that Bertrand Russell mistakenly assumes that significant sentences must be about something (Strawson, “On Referring” 323). Strawson believes Russell ignores the distinction between the use of a sentence and the sentence itself, causing Russell to develop an elaborate and ultimately unnecessary theory. In this paper, I will outline Russell’s theory, giving reasons that Russell could have had for his assumption. Then I will put forth Strawson’s criticism and his own theory. I will show that neither Russell’s nor Strawson’s theory can adequately handle all problem cases but that a combination of their views offers the first step toward a viable solution. In particular, I will argue that elements of both Strawson’s distinction between the use of a sentence and the sentence itself and Russell’s method of logical analysis are necessary for a robust theory of how sentences function.

To see the impetus for Russell’s theory, consider the following two sentences:

S1. The current wife of Prince William of Wales is patient.

S2. The current wife of George W. Bush is patient.

1 Hereafter this paper will be cited by the author’s last name followed by the number of the page on which the cited material occurs.

Jeff Johnson is a senior majoring in philosophy at Brigham Young University. After graduation in December 2006, he will attend graduate or law school. This essay won first place in the 2006 David H. Yarn Essay Contest.
Both of these sentences are significant in a way that poorly formed formulas like “box table go” are not. They also both seem more significant than sentences like “My red dream fights peacefully.” S2 indeed seems quite easy to understand: I am predicating a certain property about the First Lady, and this predication is true or false. However, S1 is more problematic. Given that it is significant and thus likely to express a proposition, what is S1’s truth value? If we gathered all of the patient individuals and all of the rash individuals into separate rooms, Prince William’s wife would be in neither room—she doesn’t exist. How, then, could a sentence like S1, that seems to be no different from S2, have a truth value, if there are no objects to which any parts of the sentence refer?

Here Russell’s assumption that significant sentences must be about something forces him into something of a dilemma: he can hold either that S1 is not a significant sentence or he can hold that S1 is about something, but not both. Clearly it is significant, so he must hold that it is about something. However, since there is no current wife of Prince William, we either have a sentence that is about a non-existing object that still (somehow) has properties, or we have to devise a way in which the sentence can be about something other than such fictional entities. As Russell notes, Meinong holds the former view; there are non-subsisting but existing round squares, wives of current bachelors, etc. (Russell, “On Denoting” 14). However, this view seems to violate the law of non-contradiction and thus to be incoherent. It would be more pleasing to have a theory that avoids such odd metaphysical entities, and it would be nice to have a theory that does not come so close to violating the law of non-contradiction on questions of existence (the difference between “existing” and “subsisting” seems quite small).

Russell, then, must find a solution that does not posit odd entities but still makes sentences like S1 about something. To do so, he distinguishes between the grammatical subject and the logical subject of a sentence (Russell 242–43). The former is simply what the sentence seems on the surface to be about; for S1, the grammatical subject is “the current wife of Prince William.” However, Russell holds that something far more

---

2 This article will be cited by its title followed by the page number.
interesting is happening with the logical subject of such sentences. Russell holds that “denoting phrases” like “a dog,” “the dog,” and “all dogs” function not through simply referring but rather by turning the sentences in which they occur into functions of different sorts (Russell 236–37). Thus, S1 does not really say anything about Prince William’s wife. Rather, in Russell’s terminology, it states “The sentence ‘One and only one thing has the property of being currently married to William and that thing has patience’ is always true.” 3 Rendered more in keeping with modern quantificational logic, we would analyze S1 as “There exists one and only one thing such that it stands in the relation of marriage to Prince William and that thing has patience.” Clearly, the offending denoting phrase (in this case “The current wife of Prince William”) has dropped out of our analyzed sentence, supposedly taking the problem it causes with it. It also is readily apparent that our new analyzed sentence can indeed be about something; in fact, “some thing” is precisely what it is about. Russell then has a solution to such sentences that avoids Meinongian extremes while allowing him to keep the assumptions that all significant sentences are about something and that sentences like S1 are significant. We are really talking about some thing in sentences like S1, though we are certainly incorrectly ascribing some of its properties.

Strawson, however, argues that Russell’s assumption has forced him into this position and that such a position is unnecessary if one reasonably abandons the assumption. Strawson argues for this by showing the distinction between the uses of sentences or expressions and the meaning of these sentences and expressions (Strawson 327). Perhaps this distinction can be seen best by considering how we would feel about S1 uttered now, while Prince William is a bachelor, and uttered soon after Prince William gets married, if ever a girl be so lucky. It seems that in the latter instance the problems vanish; this use of the sentence is clearly about Prince William’s obviously existing wife and is significant. There is, then, something about the use of a sentence that imbues it with properties that the sentence itself, stripped of all use and context, does not have. By recognizing this fact, we can allow sentences like S1 to be significant without having to find logical

3 The analysis is different for each of the different types of denoting phrases such as “a dog” and “the dog.” For brevity, I will only give the analysis for sentences with denoting phrases containing the definite article “the.”
subjects for them (Strawson 328). Instead, we can say that, because they can be used in our linguistic community, they are significant, though this in no way means that they must be about something.

Strawson then develops a more general theory that accounts for these problem sentences. He distinguishes, as mentioned above, between a sentence and a use of a sentence. He also distinguishes between these two things and an utterance of a sentence. The uses of a sentence are something like what the speaker can accomplish with that sentence. The utterance of a sentence is something like the instantiation of a sentence under a certain use at a particular time and place by a speaker. For example, I can use the sentence “I like pizza” to give my wife instructions on my preference for tonight’s dinner or to express my general gustatory preferences to a stranger. For Strawson, meaning occurs in sentences themselves. That is, sentences give “general directions” for their usage according to the rules and customs of a linguistic community (Strawson 327). Following the previous example, the general directions given by the sentence “I like pizza” are probably something like “use this sentence to inform someone about your preference for a certain kind of food,” thus making it possible for us to use this sentence in both particular situations (my wife’s question about what I want for dinner, pizza or salmon) or more generally (the stranger’s query). Truth occurs at the level of utterance and usage; only after a sentence has been uttered under a particular usage can it have a truth value. Hence, according to this more general theory, sentences like S1 can never by themselves have a truth value (Strawson 329). But the truth value (or lack thereof) of such sentences was one motivating factor leading Russell to look for the logical subject. With Strawson’s new theory in hand, we can dismiss Russell’s problem sentence and properly distinguish between the meaning and reference of sentences and expressions.

Russell, though, can immediately reply that Strawson simply points out a distinction that Russell was implicitly (and reasonably) assuming. The first reason for thinking that Russell assumes this distinction is that he describes his example sentences, sentences like S1, as “puzzles” (Russell 240). Clearly, such sentences are only puzzles when they are used in a very specific context, namely the context in which the object supposedly referred to by the denoting phrase does not exist. Russell would have no
problem with an Englishman in 1745 uttering statements about the present King of France, nor would he have problems with a contemporary Englishman making similar statements while acting in a play. The second reason to assume Russell intended to consider only declarative uses of such sentences is that it is almost always the case that when someone utters or writes a sentence, they intend to assert a proposition. Granted, in philosophical examples the author may not intend for others to think that he or she is actually asserting the sentence, but the author still intends that the reader consider the sentence as a contemporary utterance unless otherwise specified. Consequently, it is very plausible to believe that Russell intended to consider only uses of sentences, making Strawson’s distinctions irrelevant or implicitly operative in Russell’s work.

This reply, though, is somewhat unsatisfactory. If Russell’s theory already has accounted for the use of sentences (despite perhaps his sloppy use of philosophical language), then why the elaborate hunt for the logical subject? Remember, the assumption that significant sentences need to be about something led us to the view that they are really about their logical subjects. Do we still need to hunt for logical subjects of sentences when only uses of sentences are under consideration? Is there still a place for Russellian analysis if we accept the distinction? More to the point, can Russell’s analysis account for something in the use of sentences like S1 that Strawson’s cannot? I believe that it can. Additionally, a theory with Strawson’s distinction can avoid problems that Russell ignores. The Strawsonian part of the theory will allow us to dismiss as problems bare sentences like S1. The Russellian part of the theory will allow us to explain what problematic uses of sentences like S1 are about.

It is easy to see that Strawson’s distinction removes S1 as a problem sentence. If the assumption that all significant sentences are about something is removed, the premise that S1 is significant no longer entails that it be about something, allowing us to stop looking for what it is about. The fact that there is more left to be done after Strawson’s distinctions are drawn is harder to see. Consider, though, Strawson’s own memorable example of his handkerchief. He states, “If I talk about my handkerchief, I can, perhaps, produce the object I am referring to out of my pocket. I can’t produce the meaning of the expression ‘my handkerchief,’ out of my
Imagine, though, Strawson telling his wife on a particular occasion that his handkerchief is clean. Imagine also that at the time of this utterance Strawson in fact has no handkerchief. What will Strawson’s wife think about Strawson’s most recent usage of the meaningful sentence “My handkerchief is clean”? It seems that she will, with a puzzling look on her face, ask him if he is mistakenly talking about something else, perhaps his red necktie. If he insists that, no, his red handkerchief is clean, she will most likely either wonder if he has lost his ability to speak the English language or if his mental machinery is no longer functioning. Without there being some object that he is trying to talk about, his use of the sentence “My red handkerchief is clean” becomes exceptionally bizarre.

Indeed, this use of the sentence may even run the risk of becoming nonsensical or insignificant according to Strawson’s view. If sentences are meaningful because they give general directions for their use according to the rules of the linguistic community, sentences like the one above seem clearly in violation of all such rules. Indeed, this use of the sentence threatens to become just as incomprehensible, on Strawson’s theory, as the use of the sentence “My red book fights peacefully.” Although the latter has at present no possible usage in our linguistic community, it still seems to be meaningful. Moreover, any usage of it seems the same as a usage of the handkerchief sentence at a time in which there is no handkerchief. Intuitively, though, it seems that there is a difference between these sentences, a difference for which Strawson’s theory cannot account. We simply do not and should not use expressions to assert something about objects when we know or believe that such objects do not exist. If such assertions become nonsensical because they are not about anything, then Strawson’s argument is self-refuting because he can no longer explain significant uses of sentences. That is, if his theory is true, then a usage such as the one above violates all general directions laid down by the sentence and my linguistic community. How, though, could I ever use a sentence that did this if it is precisely those rules which give meaning to my sentence? Strawson’s theory cannot explain why all uses of sentences must be about something.

Unfortunately for Strawson, we have good reasons for believing that all uses of significant sentences must be about something, even ones like S1
or Strawson’s statement about his handkerchief. His theory, then, needs to be modified. I believe that a Russellian modification, one in which uses of sentences can be analyzed in a way similar to how Russell analyzes sentences independent of their use, can solve these problems. To show this, we can look at the independent reasons we have for believing that significant uses must be about something. One of these reasons is that scientific discourse and the advancement of science seem to depend on this fact. Consider for example the numerous references in historical scientific literature to “the ether” and “phlogiston.” The first of these was believed to fill the immensity of space in order to explain the propagation of light waves and other interstellar events. The latter was understood as a part of every combustible substance, the part given off in the combustion process. Phlogiston, then, explained the fact that the combusted object (usually) lost mass while the atmosphere gained mass. On Russell’s view, claims about the ether and about phlogiston are analyzable into claims about other objects, perhaps the real but as yet undiscovered causes of certain phenomena. This view seems to explain the process of scientific discovery. After a scientific theory is put forth, experiments are designed to test the theory. On Strawson’s view, though, any use of such a statement must go against the basic rule that we only talk in those ways when such objects exist. This gives us pause when thinking that such uses are actually about something since it is hard to see how we could be talking about anything if we are violating all of the rules laid down in the sentence by the linguistic community. If the statements about the ether or phlogiston are not really about some real thing in the world, then experiments could not have been designed to test those theories, something that clearly happened. Indeed, how does one go about testing something that does not exist or something to which no one actually referred?

Additionally, it is rarely the case that all of the phenomena explained by notions such as the ether and phlogiston are explained falsely; there are usually many things learned about nature from false theories. On Russell’s view, this process makes sense. Although there must be at least some properties falsely ascribed in sentences containing such notions, not all of the

---

4 Strictly speaking, “phlogiston” is not a denoting phrase of the type Russell mentions early in his article (Russell 235). However, Russell did believe that the vast majority of names were actually shortened forms of denoting phrases, and thus the comments made above are relevant to the discussion of denoting phrases and the sentences in which they occur.
properties need be falsely ascribed. If a scientist made a claim about the ether that was actually about the real cause of interstellar phenomena, it becomes much more plausible to think that the newer explanation of the phenomena will still incorporate true aspects that the sentences with the denoting phrase "the ether" could explain while eliminating the falsely ascribed properties of the ether. The existence of science and scientific discovery thus weighs strongly in favor of supposing that uses of sentences like S1 must mistakenly describe some real thing. This result, though, itself weighs in favor of a Russellian modification of Strawson’s views.

Another reason for supposing that significant sentences must be about something is the fact that normal discourse seems to require the supposition that individuals, when they are asserting, assert truthfully. Obviously, the questions of whether or not significant sentences must have truth values and whether or not they must be about something are distinct. However, they seem to be intimately related. If a sentence is true, in what sense can it be so if there is not some part of reality, some object, to which the sentence in some way relates? It is hard to see how the sentence “My handkerchief is clean,” asserted by Strawson to his wife, can be true if he owns no handkerchief. Normal discourse functions on the implicit assumption that individuals assert correctly. If, when a woman shouted "The President has been assassinated," I had to stop and wonder whether or not she intended to assert, whether or not she had asserted truly, and whether or not her assertion was about George W. Bush, the use of language would become exceptionally laborious and perhaps even impossible. The reason that it may be impossible is that if we cannot assume that people are telling the truth, how could we ever determine the answer to questions about whether a speaker intended to assert or intended to talk about a particular entity? It seems then that we have at least a practical reason, if not a fully theoretical one, for believing that uses of sentences must be about something, as we must assume that individuals are speaking the truth in their declarative sentences. Consequently, Russell’s position, one that explains what sentences are about even in breakdown situations like those exemplified by S1, seems more plausible than Strawson’s, which does not.

We have shown, then, that neither Russell’s nor Strawson’s theory is entirely adequate. Russell, by either mistakenly assuming that significant
sentences themselves must be about something and by failing to demonstrate how his analyses apply to the use of sentences, has created problems where in fact there are none. Strawson, by failing to explain the relationship between the general directions given by a sentence and uses in which those directions are violated, has made understanding the significance of uses of sentences like S1 quite difficult. However, by incorporating aspects of each theory into a newer one, these problems may be solved. A theory that allows for both Strawson’s distinction between the use of a sentence and the sentence itself and Russell’s logical analysis will be able to solve these problems. The former part of the theory will reject as problems bare sentences like S1. The latter part of the theory will enable us to understand what is going on in uses of sentences like S1 at times in which the grammatical subject does not exist. At least in this case, it seems that compromise does in fact solve the problem.\footnote{I would like to thank Professor David Jensen of Brigham Young University and Joshua Roberts-Gillon for helpful discussions about this and related topics in the philosophy of language. I would also like to thank Tully Minoski and Russell Farr of Aporia for helpful revisions of and responses to this paper.}

In his book *Ethics: Inventing Right and Wrong*, J. L. Mackie argues against the existence of objective moral values. He does so in two sections, the first called “the argument from relativity” and the second “the argument from queerness” (Mackie, *Ethics: Inventing Right and Wrong* 36, 38). In this paper, I will examine each argument in turn and show that Mackie fails in his attempts. Of course, by showing that Mackie’s arguments fail, I have not shown that objective values exist, but that is not the purpose of this paper. I only wish to expose some problems in Mackie’s arguments. That is, in this paper, I plan to show that the arguments from relativity and queerness are not convincing. In doing so, I will take some time to formulate them in the strongest way that I can, for I wish to avoid a straw man. Again, I am not making an appeal to ignorance; I realize that there may be other arguments better than Mackie’s and that I am far from establishing the existence of objective moral values. I am showing, though, that Mackie’s efforts fail.

The first argument that Mackie makes is from relativity. The argument from relativity has two parts, which I will call “A” and “B.” Part A takes as evidence that there are variations in moral codes between different periods of time, different groups, and even disagreement within groups (Mackie 36). The argument claims that if there were objective moral values,

---

1 This book will be cited simply by the author’s last name followed by the page number on which the cited material occurs.

---

Russell Farr is a junior majoring in philosophy at Brigham Young University. Upon graduation he plans to attend law school. This essay won second place in the 2006 David H. Yarn Essay Contest.
there would be some consensus on morality because moral systems would be based on those values. But there is widespread disagreement about morality, and this is inconsistent with the hypothesis. Therefore, we must admit that there are no objective moral values. More succinctly, the argument is as follows:

(1) If there were objective moral values, then there would be uniformity in moral codes.

(2) Moral codes differ in important ways.

(3) Therefore, there are no objective moral values.

One may object to this argument by showing there is disagreement in scientific questions, and thus one could analogously say:

(4) If there were objective facts, then scientists would agree about them.

(5) Scientists disagree about many things.

(6) Therefore, there are no objective facts.

But unless we want to adhere to extreme subjectivism, which is not a doctrine that Mackie is espousing (see Brink, “Moral Realism and the Skeptical Arguments from Disagreement and Queerness” 112), we must admit that

(7) Science is discovering objective facts.

So how is it that we can disagree about scientific questions and not end up with this contradiction? It is because scientific disagreements are the result of hypotheses that are based on inadequate evidence (Mackie 36). Our incomplete knowledge causes these disputes, and because we do not have a complete understanding of the natural world, we cannot assume that we would agree upon all objective facts. It seems then that Mackie rejects (4) and avoids the contradiction by claiming that the existence of objective facts
does not necessitate our knowledge of them. But Mackie says “it is hardly plausible to interpret moral disagreement in the same way” (Mackie 36).

In part B, Mackie makes an argument to best explanation and proposes a theory of how morality actually works. He claims that our morality is a result of our habits and customs (Mackie 36). For example, we prefer monogamy because it is the type of thing we do, not because monogamy is an objective good. We idealize the things that we do and call them objective values. This theory better explains the variety in moral codes (Mackie 37). If people in different cultures have different moral values than we do, it is because they have idealized their behavior, some of which is different than our own. This theory can adequately explain the wide variety in moral codes, a variety that causes problems for theories positing objective moral values.²

The question, though, is whether or not Mackie’s arguments are actually convincing. If he is trying to demonstrate that there are no objective moral values, he has a difficult task. There are very few ways of proving that something does not exist. Either we would have to discover some true conditional by which we can use *modus ponens* to arrive at the non-existence of the thing or we would need to resort to a *reductio ad absurdum*. Mackie seems to propose the conditional “if there are objective moral values, there would be a consensus” and infers there are no objective moral values by *modus tollens*.

It seems Mackie has made a fatal assumption here. He has assumed that the existence of objective moral values would force a consensus. I fail to see why that would be so. What is it about objective moral values that would force their knowledge upon us? There are many objective truths that are yet unknown or are disputed by us. For a long time scientists have debated the makeup of the atom. It cannot be both Thompson’s plum pudding and an electron orbiting a nucleus.³ Yet whatever is really the makeup of the atom, various possibilities have been disputed. Objective facts have yet to force their knowledge upon us in any other circumstances, so why should morality be different?

Mackie does try to refute objections that compare moral disagreement to scientific disagreement, saying that “it is hardly plausible to

---

² Mackie treats a theory by Sidgwick that proposes to overcome the difficulty while maintaining objective values (Mackie 37), but this is not relevant to my discussion.

³ The current model is an electron probability cloud surrounding a nucleus.
interpret moral disagreement in the same way” (Mackie 36). But he gives no other reason to believe this than his assertion. Scientific pursuits can avoid this problem because the disputes are based on inadequate knowledge. Yet moral disagreements could come about the same way. The key to Mackie’s refutation of the counterexample is knowledge. Because objective facts do not cause us to know them, we must search for knowledge; in that search we often err. But, in order to conclude that moral systems cannot avoid this problem in the same way scientific disputes do, we must assume that all people know these objective moral values equally well. This assumption has little in the way of justification.

If there are objective moral values, we have no reason to suppose that all people know them equally well, and so we can expect some disagreements about morality. These disagreements would be based on one or both disagreeing parties’ faulty understanding of what the true objective moral values are. To give an example with a strongly moral aspect, let us look at the Qur’an. The Qur’an is a book containing codes of conduct that a faithful Muslim must observe, yet those who call themselves devout Muslims range from the peaceful Sufi mystics to the Jihadists. Both of these groups read the same book and get wildly different meanings from it. So, even when the law is written down and easily accessible to all, people will interpret it differently. If we cannot agree upon moral codes that are written for all to see, I do not see how we are justified in assuming that everyone must know and understand an unwritten moral code equally well. Therefore, we can reject (1) and can thereby demonstrate that Mackie has not adequately supported (3).

Even supposing that there was an objective value that was known equally well by two parties and that those parties in some way disagreed about it, this would only show that one of the parties failed to implement it correctly or that this particular supposedly objective value indeed does not exist. Disagreement cannot show that the whole set of objective moral values does not exist, only that one or more value supposedly in the set does not exist. Such is the fate of arguments of this type, for example, the related arguments dealing with the problem of evil. If they are sound, they do not show that God does not exist but that a being with all the properties of the Absolute cannot exist. There is a possibility that Mackie’s argument
could show that some certain moral value or values do not exist, but even this allows too much. Disagreement as it stands may be explained by human error in our efforts to understand exactly what the objective moral values really are.

Richard T. Garner asserts that Mackie makes no attempt to show that objective values do not exist by such an argument as the one refuted above; instead, he claims that Mackie’s argument is an argument to best explanation (Garner, “On the Genuine Queerness of Moral Properties and Facts” 140). I will now examine this version, which I have previously called “B,” but I should note that even if Mackie’s theory best explained our moral systems, it would not show that there are no objective values; it would only show that those are not the sort of thing that all people base their morality on. I say “all” because this disagreement does not exclude the possibility that some groups may base their morality on objective values.

This version of the argument makes the assumption that there is vast disagreement among moral codes, this being the anomaly that Mackie’s moral theory explains better. If there were no vast disagreement, then the subjectivist would have to explain how so many societies came to behave in such incredibly similar ways and then idealize those behaviors. Merely mild disagreement does not show that there are no objective moral values because objective moral values would help explain the almost overwhelming consensus of the world’s people. The disagreement among peoples must be vast enough that it actually counts against the objectivity of values. But those who make this argument have severely overestimated the level of disagreement in the world.

Even our moral disagreements rest on a foundation of broad moral agreement. For example, in the ongoing debate about abortion no one questions whether it is right or wrong to kill people. Pro-choice advocates do not condone murder. The questions they raise are about whether a fetus is a person with a right to life. In other disagreements, only some aspect of an objective moral value is questioned. For example, death penalty proponents and opponents do not disagree about whether killing people is morally wrong; the issue is whether it is morally justified, or even required, in the case of extreme criminals. Even Jihadists appeal to values with which we can identify. They indiscriminately kill because they truly believe the
West is trying to destroy their people, culture, and religion. We disagree with their assessment and their methods, but we do not disagree that people should protect their people, culture, and religion. We may argue with them to try and convince them that non-violent methods are more appropriate, but we can identify with their love of their own society. Even sadists appeal to values we can understand. They inflict pain on others because it gives them pleasure. We understand why a sadist values pleasure because we value it ourselves. We non-sadists simply disagree with how the person achieves such pleasures and would say that the sadist values pleasure above more important values. We can see that the questions in many of these disagreements are either about facts or about which moral values have preeminence, not about two utterly foreign moral values.

That we can even communicate about moral values with people whose moral codes clash with our own shows some consensus. Those who have differing moral codes will always justify their practices with a value we can understand. Once we realize this, Mackie’s claim to have the better explanation falters. Did people from separate communities all across the world independently happen to behave in ways that led them to idealize nearly all of the same values that all other peoples also idealized? Or is this agreement explained better by the theory that there is something about these values that draw all peoples? The disagreements are too few and the consensus is too wide for us to easily accept Mackie’s theory.

Mackie also argues against objective moral values by discussing their “queerness,” claiming that objective moral values would be unacceptably queer both metaphysically and epistemically (Mackie 38). Whatever kinds of things objective moral values are, they would have to be very strange entities, so strange that we could not bear it. His metaphysical argument seems to be:

(8) If there were objective values, they would be unacceptably queer.

(9) Whatever would be unacceptably queer cannot exist.

From lines 8 and 9 we can deduce:

(10) There are no objective values.
The epistemic argument claims that coming to know what objective moral values are and understanding their authoritative prescriptivity would require some special faculty. Whatever sensory perception is necessary to know these objective values would be unacceptably queer. Thus, we can conclude that we cannot know objective values by an argument similar to the one above.

Now we ask whether Mackie argues that there are no objective values at all or that there simply are no objective moral values. If he is claiming that there are no objective values whatsoever, we must ask by what standard he can judge moral values as strange. If there were no objective values, such as aesthetic values, then any standard by which we judge moral values would have to be completely subjective. And we should ask why we must choose the subjective standards that label moral values unacceptably queer. Just as we can freely choose any moral values to live by if there are no objective values, we can freely choose the standards by which we judge the queerness of objective moral values. And those standards might not label moral values as queer. In fact, we could claim that objective moral values are the only things in the universe that are not unacceptably queer, or that a universe without objective moral values is unacceptably queer. Therefore, if there are no objective values at all, we can claim objective moral values are not unacceptably queer.

If Mackie claims that there are some objective values, namely aesthetic values, and by these values we could objectively say moral values are too queer, then we could propose that objective moral values are of the same kind as the aesthetic values. Moral values would be of a similar type to the other objective values, and we could know moral values through the same faculty by which we know the other objective values. Therefore, if there are some objective values, we can claim that objective moral values are not unacceptably queer. And because the law of excluded middle dictates that either “there are no objective values” or “there are at least some object values” is true, we can claim that objective moral values are not unacceptably queer. That is:

(11) If there are no objective values, objective moral values are not unacceptably queer.
(12) If there are at least some objective values, then objective moral values are not unacceptably queer.

(13) There are either no objective values or there are at least some objective values.

(14) Therefore, objective moral values are not unacceptably queer.

And so, even if we grant Mackie’s assumption that nothing unacceptably queer exists, we need not conclude that objective moral values do not exist.

Finally, that Mackie finds objective values too strange for comfort seems an odd argument against their existence. I find electrons to be very strange. The idea that something can be both a wave and a particle and that it exists as a definite cloud of probability until a measurement collapses its wave function is a very strange notion indeed! If I find some idea strange, does that count in any way against its truth? Certainly not! Moreover, Mackie finds objective values unacceptably strange because he construes them so. He claims that objective values would have to be Platonic Forms—such Platonic Forms would be strange to be sure, but they are not what many objectivists claim these values are.

When we are confronted with arguments of this type, we need to examine the assumptions underlying them. Mackie’s assumptions, that objective moral values would force a consensus and that objective moral values would be unacceptably queer, are unfounded. Thus, while I have not shown that there are objective morals, I have shown that Mackie’s arguments from relativity and queerness fail to persuade us that objective moral values do not exist.
Works Cited


In “Reference and Definite Descriptions,” Keith Donnellan makes a distinction between attributive and referential uses of definite descriptions. Donnellan argues that Russell’s theory of definite descriptions does not accommodate this distinction. Saul Kripke, in “Speaker’s Reference and Semantic Reference,” argues that although Donnellan’s distinctions may have some pragmatic import, they do not refute Russell’s theory. Donnellan challenges Russell’s truth values for denoting sentences, claiming that they cannot correctly deal with his referential use. He proposes, against Kripke’s claim, that there is some semantic relevance to his distinction. Is Donnellan’s distinction relevant to semantics and truth conditions or not? In this paper I will begin by summarizing Donnellan’s distinctions and then will show how they make Russell’s theory inadequate. I will then discuss Kripke and Donnellan’s arguments and distinctions, concluding with the questions of whether or not belief affects Donnellan’s distinctions and whether or not his distinctions are solely pragmatic.

Donnellan distinguishes between attributive and referential uses of definite descriptions. The former occurs when the speaker predicates a property of whatever is referred to and the latter occurs when the speaker gives his audience the ability to discern whom or what he is talking about. Donnellan uses the example “Smith’s murderer is insane.” Suppose we find Smith’s body eviscerated in a brutal fashion, and I comment,

Alan Feuerlein is a junior majoring in philosophy at Brigham Young University. After graduation, he plans on attending law school or studying mathematical logic. This essay took third place in the 2006 David H. Yarn Essay Contest.
“Smith’s murderer is insane.” This use is attributive; I predicate insanity of Smith’s murderer, whoever it is. Suppose that at the murder trial, the defendant Patrick is spinning in circles and babbling to himself. I comment, “Smith’s murderer is insane.” When asked whom I’m speaking of, the answer is Patrick. So, using Donnellan’s distinction, this use is referential. Assume no one murdered Smith. If the definite description is used attributively, then no one is falsely said to be insane because we attempted to attribute something to someone who does not exist. Conversely, if the definite description is used referentially, then Patrick is still said to be insane because we are referring to Patrick with the definite description “Smith’s murderer.” An attributive use occurs whenever we presuppose that something fits the definite description. Attributively we suppose that something or other (even if it is unknown) fits the definite description, and referentially we suppose that a particular fits it.

Thus, when we speak attributively the description is essential. If we want to predicate something of “Smith’s murderer,” it is necessary that we use the description—otherwise, meaning is lost. But referentially, any description would have worked, since we are referring to a certain object or person. For example, if speaking referentially, we could have just as easily have said, “Patrick is insane” or “That man who is spinning is insane.” So a question arises: is the difference between the referential and attributive use just a difference in the belief of the speaker? Donnellan answers no. He states, “It is possible for a definite description to be used attributively even though the speaker (and his audience) believes that a certain person or thing fits the description. And it is possible for a definite description to be used referentially where the speaker believes nothing fits the description” (Donnellan, “Reference and Definite Description” 251). Suppose my mother is married to a man who I am told is my father, but I believe he is not. If my mother is on the phone with him, I may say, “May I speak with my father?” I succeed in referring to him without me or my mother being forced to believe he is my father. I use the definite description “my father” referentially without any corresponding belief. So the question is whether or not Donnellan’s distinction is semantically relevant or merely pragmatic.

---

1 This work will be referred to by the author’s last name followed by the page number.
Donnellan challenges Russell’s view of the truth values of referential uses. He proposes that there is some semantic relevance to his distinction. According to Donnellan, even in speech, if nothing fits the description (for example, if Smith committed suicide), then “it is [still] possible to say something true or to ask a question that gets answered or to issue a command that gets obeyed” (Donnellan 252). The problem Donnellan has with Russell’s theory can be best understood by example. As in the earlier example, even if Smith committed suicide, when I am in the courtroom and refer to Patrick by saying “Smith’s murderer is insane,” what I said is true if Patrick is in fact insane since I am using the definite description referentially. The problem for Russell, then, is how his theory analyzes the referential use of this statement. Russell’s theory, when applied to the referential use in which \( x \) serves as the description, entails that “there exists one and only one \( x \).” Donnellan states that “The implication that something is the \([x]\) . . . does not amount to an entailment” (Donnellan 253). So even though when we use the phrase “Smith’s murderer is insane” referentially we may imply that there is a murderer, it does not entail or specifically state that there is such a thing. According to Russell’s theory, if Smith committed suicide, the phrase is false, since there is no unique \( x \). But it seems clear that when we use the phrase we successfully refer to Patrick, and that, if he is insane, the statement is true.

“One might think that if Donnellan is right, Russell must be wrong, since Donnellan’s truth conditions for statements containing referential definite descriptions differ from Russell’s” (Kripke, “Speaker’s Reference and Semantic Reference” 253).² According to Kripke, though, there is no argument against Russell here. Donnellan could be right about his distinction, but only in a pragmatic sense, and Russell could likewise be correct about a statement’s semantics or truth values. Kripke considers the example “Her husband is kind to her,” used when the woman has no husband, only a lover. Kripke asserts that in this case, Donnellan would say we referred to the lover, and said of him that he is kind to her. Kripke questions whether Donnellan would say that this statement is true or false; he says of this that “Donnellan would hedge” (“Speaker’s Reference and

² This paper will be referred to by its title followed by the number of the page on which the citation occurs. Kripke’s Naming and Necessity will be referred to by its title followed by the page number as well.
Semantic Reference” 253). Kripke points out that if Donnellan were to say the statement would be true if the lover is kind to her, then there would be a point of contention between him and Russell, since Russell would assert that the statement is false. But according to Kripke, Donnellan is unclear on this. Kripke states that in this case, when reporting the statement, we must use the definite description “her husband” either attributively or referentially. If it is attributively then “we misrepresent the linguistic performance of the speaker” (“Speaker’s Reference and Semantic Reference” 253). Basically we would be saying something like her husband, whoever he is, is nice to her, when the fact is that the man is speaking of the lover and not her husband, whether or not she has one. So we turn to the referential use. If speaking referentially, we are referring to the man there, but generally we refer to someone as “her husband” or “his wife” only if it is assumed they are actually married. Kripke claims that “Donnellan does not clearly assert that the statement . . . ever has non-Russellian truth conditions” (“Speaker’s Reference and Semantic Reference” 254). But Donnellan states that if we do use the definite description “her husband” referentially then “we are ourselves referring to someone and reporting the speaker to have said something of that person, in which case we are back to the possibility the he did say something true or false of that person” (Donnellan 257). Donnellan concludes that when using a definite description, we may be saying something true or false, but he gives no clear indication that there are cases in which an uttered statement may be neither true nor false.

So who is right in this case, Kripke or Donnellan? Is there is a semantic distinction to be made when dealing with definite descriptions, or is Donnellan’s distinction solely a pragmatic one? In Naming and Necessity, Kripke states that “[Donnellan’s] remarks about reference have little to do with semantics or truth conditions, though they may be relevant to a theory of speech-acts” (Naming and Necessity 25). Donnellan on the other hand holds that when we use a definite description referentially, even if nothing fits the description, we “may have stated something true or false” (“Speaker’s Reference and Semantic Reference” 257). It seems that much of the apparent controversy may have to do with the belief of the speaker. So we are back to the question Donnellan asked of his own theory previously, whether the differences between the referential and attributive use just
come down to a difference in the belief of the speaker. If it is simply a matter of belief, then Donnellan’s distinction has little to do with semantics. If not, then Donnellan’s distinction is at odds with Russell and is of more important consideration in theories of definite descriptions.

Donnellan admits that when we say something attributively using the form “The \(x\) is \(F\),” where there is no \(x\), we can’t report whether or not the speaker said of this or that person that he or she is \(F\). But if we say something referentially then we can attribute something to \(x\). Take the example of “Her husband is kind to her.” Assume the that the man the speaker was referring to is Jones. In this case, we may correctly report that “Jones is kind to her” or “The \(x\) is \(F\),” where \(x\) is “Jones” and \(F\) is “being kind to her.” Donnellan continues and asks us to suppose that Jones is the president of a university. We may then say, “The president of the university was kind to her” and do so correctly. If we then confront Jones himself and say, “He said that you were kind to her,” we may do so without mentioning at all that he was originally referred to as her husband. The statement still has a truth value, and no belief of Jones being her husband is implied. Kripke thinks that Donnellan concludes, “Thus where the definite description is used referentially, but does not fit what was referred to, we can report what a speaker said and agree with him by using a description or name which does fit” (“Speaker’s Reference and Semantic Reference” 256).

Kripke advocates a different view for referents of names and descriptions, which he calls the “semantic referent.” For names, this semantic referent is what is named; for descriptions, it is what uniquely satisfies the description. In this way Kripke states that a speaker may refer to something other than his semantic referent if he has appropriate false beliefs (Naming and Necessity 25). This is what Kripke really thinks is happening in some of Donnellan’s examples. If we consider the example “The man over there playing poker is sad tonight,” then according to Donnellan we can refer to him referentially even if he is not playing poker, but according to Kripke, we may only refer to him (assuming he is not playing poker) if we have appropriate false beliefs. Otherwise, we must be referring to the

---

3 Even if there is implication, there is no entailment (see above).
semantic referent, a man that is really playing poker and is sad. So the man
the speaker refers to may not be playing poker and is still sad, but there
could be another man at the party who is playing poker and is not sad.
Thus we may say that the statement is false, but we know that no one is
referring to the second man who is playing poker and not sad, even
though the definite description may semantically refer to him.

To summarize, Kripke seems to think that Donnellan’s distinction
has pragmatic rather than semantic import; it does not refute Russell’s the-
ory, but only gives a description of how definite descriptions are used in
language. Donnellan, on the other hand, seems to have shown that his dis-
tinctions are semantically important, that they affect truth values, and that
the interpretation of such statements is necessarily different than Russell’s
in a way for which Russell’s theory cannot account.

It seems that even if we do believe that there is something else which
fills our definite description we may still refer to some object other than
that which semantically (in Kripke’s sense) fills our definite description. So
in Donnellan’s sense, if the man referred to really is sad “we have a sub-
stantial intuition that the speaker said something true of the man to whom
he referred in spite of his misimpression” (“Speaker’s Reference and
Semantic Reference” 249). It seems to me that Donnellan’s distinction can-
not be simply reduced to a question of belief. Whether or not I believe I am
referring to the correct man, or if it is unknown to me whether there is
another man who fits my definite description, I can still have an attributive
and referential use, and apply a truth value to such uses. For example, con-
sider the statement “The red book there is long.” Now if for example I
read this book and then utter this statement, I may be using it attributively,
stating that the book is long. For an example of a referential use, someone
may ask me, “Do you have a long book?” I can respond referentially and
say, “The red book there is long,” referring to the red book. Now in this
example suppose that the book is really orange, but in a certain light it
looks red to me, then I can still refer to the book by stating what I said,
even though it may not really be the case. The item I referred to, even
though it was orange and not red, is still long, and since I successfully
referred to it, my statement was true. The same applies for the attributive
use. Also, even if I believe the book to be orange but know that it is often
mistaken for a red book, I can use the statement “The red book is long” either attributively or referentially and still apply truth conditions to it.

It seems as though the truth of a statement comes after its interpretation, not in the immediate utterance. If I say, “The red book is long,” even though it is really orange, if I do an immediate evaluation of the book without questioning beliefs, then the statement is false. The book is orange, not red. But if we interpret the statement after we have used it either attributively or referentially, the statement is true if we correctly refer and there is an audience. It seems there may be an ambiguity in this, but I think that it depends partially on when our statement is interpreted. Beliefs and interpretation are thus important in the sense of being able to correctly assign a truth value, for if no beliefs are considered, then the intention of the speaker can easily be lost.

So to conclude, beliefs seem not to impact Donnellan’s theory so long as we consider the interpretation of a statement, not just the statement itself. Truth conditions may still be applied to definite descriptions via Donnellan’s distinction. Kripke’s idea of the semantic referent may work as well, but he has not shown fully that Donnellan’s theory does not have semantic import and or that it has no effect on Russell’s theory of denoting phrases.

4 It also seems that even though the book may be orange and not red, the main property that I am predicating of the book is that it is long, so we are most concerned with this. In some sense I suppose this could be considered false since the book is orange, but the spoken statement still seems true; it is strange to just think of it as empty or false since it is obviously neither of these.
Works Cited


Ever since Morris Weitz in his classic 1956 paper “The Role of Theory in Aesthetics” condemned the classical sense of defining art in essentialist terms, philosophers, aestheticians, and art critics alike have increasingly concentrated upon limit cases. Weitz, influenced by Wittgenstein,1 presented an argument that pointed out the shortcomings of popular definitions of art. These definitions failed because they either neglected to capture features common to all works of art, neglected to capture features particular to certain artworks, or misconstrued the logic of the concept of art (Weitz, “The Role of Theory in Aesthetics”).2 Scholars were left trying to deduce a non-emotivist definition of art that was neither mimetic (as in Plato), formal (as in Clive Bell), or expressive (as in Leo Tolstoy). Furthermore, according to Weitz, any new definition should remain “open” in that it must allow for creative novelties (Weitz 189).

Now, what such creative novelties are we talking about? Well, to give an example, in the early twentieth century the expression of art’s limits became commonplace within the art world. Artists such as Marcel Duchamp, Richard Hamilton, Jasper Johns, and Andy Warhol introduced artworks that pushed the envelope from within. Their idea: “let’s produce

1 Wittgenstein argued that essentialism (the Platonic idea that words are relations to the real world, corresponding one-to-one) fails in that many words simply present family relations, not one simple uniform thread (Wittgenstein, Philosophical Investigations 65–75).

2 This work will be referred to by the author’s last name followed by the page number on which the cited material occurs.

Ryan Dreveskracht recently graduated from the University of Washington, where he majored in philosophy and law, society, and justice. In the fall of 2006 he will pursue a joint JD/PhD in Philosophy at an as-yet undecided university.
works in some procedure [and see] whether art succeeds in incorporating it.” Interestingly enough, Weitz’s theory succeeded in incorporating these revolutionary artworks.\(^3\) However, Weitz’s theory was only successful because it set no limits at all (Mandelbaum, “Family Resemblances and Generalizations Concerning the Arts”). Thus, much current aesthetic theory has reverted back to its essentialist roots, albeit attempting to arrive at a definition of art that defeats the proposed errors in essentialism by maintaining that definitions do not exist within essences. Anti-anti-essentialism occupies the middle ground, standing somewhere in between essentialism and anti-essentialism because it defines art while maintaining an “open” conception thereof. In response to anti-anti-essentialism, definitions of art such as Dickie’s institutional theory, Gaut’s cluster concept, Eaton’s culturally advanced notion, and, most significantly for this paper, Levinson’s historical definition have been proposed.\(^4\) In what follows, I will both present an analysis of Levinson’s historical definition of art and will assess the relevant objections. This analysis will be threefold since I will be considering three different types of objections to Levinson: those regarding the \(\uparrow\)-arts, the implausibility of an intrinsic recursive definition, and the disadvantages of an intentionalist-historicist account.\(^5\) In culmination, I will argue that, although Levinson’s historical definition of art seems to defeat many of the objections that have been raised against it and is not irreparable, his definition ultimately fails to address a contradiction within his intentionalist-historicist framework, specifically that between a proprietary right and the liberalized version of regard-as-a-work-of-art.

So, to begin, let’s give a brief account of Levinson’s historical definition and how it came about. Levinson’s historical definition of art, in fact, actually stemmed from an objection that he had to George Dickie’s institutional theory of art (“Defining Art Historically” 213–33).\(^6\) Roughly, Dickie’s institutional theory maintained that relations to the art world “are

---

\(^3\) Note that I have not given Weitz’s theory. It is irrelevant to the point that I am making, namely that Weitz’s theory succeeded in assimilating revolutionary artworks, artworks generally not easily assimilated.

\(^4\) Levinson notes that he now believes that his definition should have been denominated “retrospectiveist” or “auto-referentialist” (rather than “historicist”), as to avoid associations with robust dialectical-synthetic historicist accounts such as those of Hegel and Danto.

\(^5\) Although what I offer here is by no means a comprehensive account of all of the objections made against Levinson, they are the ones most detrimental to his historical definition.

\(^6\) This work and all others by Levinson will be referred to by title and page number.
the criteria by which art is defined.” But Levinson objected on two accounts. First, Levinson objected to the alleged necessity of an artistic cultural performance. Art, according to Levinson, can be private and therefore depends upon no one but the artist and the potential viewer’s regard (“Defining Art Historically” 225). According to Dickie’s theory, neither a child’s drawing nor his mother’s macramé can be considered art because they were not presented to an artistic institution. Hence, Dickie’s theory is over-restrictive. Second, Levinson presents a reductio ad absurdum. If the presentation of artwork is what defines artwork and if it is up to the art world to present the piece, which depends upon the art world’s appreciation of that piece, is not the institutional theory just a dressed-up way of equating artwork with the appreciation thereof (“Defining Art Historically” 225)? Thus, the implications of Dickie’s theory fall victim to the same limiting problems that destroyed essentialism.

However, Dickie’s theory did contain the framework from which Levinson’s hypothesis began: the definition of art in terms of non-exhibited properties, defined relationally (“Defining Art Historically” 224). Specifically, Levinson’s move away from Dickie took shape from Levinson’s intent to present artwork that is historically defined, rather than artwork that is presented to an “art world.” Furthermore, what specifies the relation is that it is intended to be presented within a historical context rather than the physical act of presentation to any establishment (“Defining Art Historically” 225). According to Levinson’s formal definition:

(I) x is an artwork if and only if x is an object that a person or persons, having the appropriate proprietary right over x, non-passingly intend for regard-as-a-work-of-art, i.e., regard in any way (or ways) in which prior artworks are or were standardly regarded (“Defining Art Historically” 227).

Notice that the definition remains “open” by having no intrinsic essentialist criteria, but it is limited in that it is exclusive to a historical connotation (“Defining Art Historically” 226). Levinson’s theory initially surpasses his objections to Dickie. For example, suppose a caveman time-traveled to our current times. Would he regard our architecture, our paintings, or

7 Note, however, that although regard-as-a-work-of-art may entail an exhibit to a third party to qualify as art, Levinson’s theory is not dependent upon some type of cultural performance or presentation.
Warhol’s Brillo Box as works of art? Certainly not! The term would have no meaning to him. Art without its historical connotation is meaningless because art depends upon both the existence of earlier art and the traditions of regarding that earlier art in particular ways.

But then two obvious questions arise. First, who qualifies as an art-aware maker? Second, since those most knowledgeable of art history are most commonly members of an artistic institution, how does this definition differ from Dickie’s? Levinson answers that art-aware makers are simply those individuals who connect their creation to what art has been; if they are not doing this, they are not consciously producing art (“Defining Art Historically” 227). This does not, however, limit our definition to the fully art-knowledgeable, for even a child painting in a kindergarten classroom knows enough about the historical content of art to intend for his painting to be regarded as an artwork as opposed to an article of clothing (“Art Historically Defined: Reply to Oppy” 155). So, Levinson defines “intends for” as that which “makes, appropriates, or concedes for the purpose of” in order to comprehend fashioned, found, and conceptual art (“Defining Art Historically” 228). Furthermore, this intent must be static rather than transient: if someone were to say, “I want to regard x as a piece of art,” x must be in the person’s possession, mentally and physically, for a sufficient amount of time before regard-for-x becomes effective (“Defining Art Historically” 228).

Also, one must regard past art standardly rather than commonly or passingly (“Defining Art Historically” 228). Although Levinson gives no specific criteria for regarding artworks correctly, he does cite five relevant considerations: how the artist intended his work to be regarded; what manner of regard is most rewarding; kinds of regard similar objects have enjoyed; the optimum regard for realizing the ends which the artist had in

---

8 For example, if our caveman were to throw a bucket of paint onto his bed, the product would not be art. This is because the caveman, not being aware of producing art as it has traditionally been regarded, could not have intended it to be art.

9 In saying “effective,” Levinson means “serious and stable” (“Extending Art Historically” 417). Levinson mentions this only in passing, but what I gather from the stipulation is that one must really mean to regard something. In other words, a man passing by his kitchen counter cannot simply point to the object and state that he intends to regard it as art; he must be sincere.

10 The word “passingly” is a technical term used by Levinson. However, it does not carry much relevance within our current discussion.
mind for appreciation; and what makes for the best development of art (“Defining Art Historically” 228).

But Levinson continues his discussion by adding that the definition includes a proprietary-right condition, meaning that one cannot regard-as-a-work-of-art an object that one does not own (“Defining Art Historically” 228). This is not to be confused, however, with objects that the artist points to or represents, such as occurs in the various conceptual arts (“Defining Art Historically” 228). Furthermore, curators, promoters, and exhibitors cannot turn non-art objects of the past into art (“Defining Art Historically” 229). For example, suppose a caveman, Laurence, brings with him a piece of finely decorated loincloth. As Laurence no longer needs his old loincloth, an art gallery promoter presents it as a piece of art. But Laurence demands the removal of his loincloth from the art gallery walls, claiming that it is a spiritual object not meant for public admiration. We are left with a dilemma. How can an object mounted for regard-as-a-work-of-art simply revert back into being non-art? Simply put, it does not. “It never was art at all, because our present art establishment unknowingly lacked the [proprietary] right to make it such” (“Defining Art Historically” 229). However, this proprietary-right condition does have one exception. Owners unopposed to the intention may grant someone permission to deem their possessions artworks (“Defining Art Historically” 228). This is essentially what happens when an artist is commissioned to create a work of art. Although the sculptor’s marble is not her possession, the owner willed it to be granted to her.

Note also that, for an artist to intend for regard-as-a-work-of-art, the intent must meet one of three criteria:

1. **Specific art-conscious intention**: The intent to regard-as-a-work-of-art in the specific way in which some particular artworks or class of artworks have been correctly regarded.

2. **Non-specific art conscious intention**: The intent to regard artwork as past artworks have been correctly regarded, having no specific ones in mind.
3. Art-unconscious intention: The intent to regard in some specific way $x$, characterized in terms of intrinsic features, where $x$ is unconsciously a way in which art has correctly been regarded ("Defining Art Historically" 229).

According to these criteria, one may intend either referentially (1 and 2) or transparently (3), allowing for conscious intention or naïveté ("Defining Art Historically" 229). Although Levinson’s idea of art-unconscious intention may seem complex, it is really quite simple. Say, for instance, that Laurence, realizing that he was in a strange world and needing to represent the strange futuristic objects that he had been exposed to, began painting mimetic pictures. In this case, since Laurence’s intent for regard was one in which art has historically been regarded, he has unconsciously created art.

But now we ask the question, what about time? Does Levinson’s theory carry time-dependent qualifications? In brief, on Levinson’s theory, “an object can be an artwork at one time and not another” ("Defining Art Historically" 230). This may happen in one of two ways. First, an object may not be an artwork at its creation but may become one at a later date, specifically after the artist has added his intent for regard-as-a-work-of-art into the work’s constitution ("Defining Art Historically" 230). An example is Duchamp’s In Advance of the Broken Arm. The shovel had previously existed in a hardware store amongst similar snow shovels before Duchamp had intended it for regard-as-a-work-of-art. However, once he bought the object and intended it to be regarded in a specific way, it became art, even though the physical composition of the object did not change at all.

Second, an object may become art after its creation if the history of art catches up with what the object’s creator was involved in ("Defining Art Historically" 231). In other words, $x$ may become art because it was intended to be regarded in a way $y$ that is one way in which we presently regard artworks, though $y$ may not have been a way in which art was correctly regarded at the time of $x$’s creation. For example, a caveman’s mimetic paintings may not have been considered artworks in 800 BC, just as commonplace roadmaps are not considered artworks today. Nonetheless, because intent for regard as mimesis is a subset of where our regard-as-a-work-of-art has evolved since 800 BC, the caveman’s mimetic
paintings would become art if anthropologists uncovered them today (“Art Historically Defined: Reply to Oppy” 154). Levinson also asserts that what art becomes through a historical evolution “depends conceptually, not just causally, on what art has been” (“Defining Art Historically” 234). This is part of the reflexive nature of the theory as it is construed upon intent rather than intrinsic or extrinsic qualities of specific artworks. So, working recursively, one may expect to discover one type of intent that was regarded as art from which all art stemmed, the so-called “ur-arts” (“Defining Art Historically” 234). According to Levinson’s definition, an object becomes an artwork when its intent is related recursively to the purposes of some ur-art of our tradition and when it descends via intentional relations which invoke previous correct regards. As Levinson says:

(II) Initial Step: Objects of the ur-arts are artworks at $t_0$ and thereafter (let $t_0$ be the time roughly at which the ur-arts began spawning the non-ur-art artworks). Recursive Step: If $x$ is an artwork prior to $t$, then $y$ is an artwork at $t$ if it is true at $t$ that some person or persons, having the appropriate proprietary right over $y$, non-passingly intends (or intended), $y$ for regard in any way (or ways) in which $x$ was properly regarded (“Defining Art Historically” 235).

But, again, this brings up a fairly obvious question: how does a reflexive definition of art allow for revolutionary artworks, artworks intended for treatment in a manner distinct from their predecessors (“Defining Art Historically” 233)? Levinson presents two methods that may accomplish this. First, for artists to make “art” despite the revolutionary aspect desired, they must initially direct their audience relative to their work in a way that artworks have previously been regarded (“Defining Art Historically” 233). For example, some, in fact most, artistic movements are revolutionary in a weak sense; impressionist paintings were meant to be regarded in some new specific aspect that was previously not a part of our history of regard-as-a-work-of-art (“Defining Art Historically” 232). However, some contemporary movements (such as Dadaism) which attempt to deny the applicability of all past ways of taking art objects require a second method, a method called the “liberalized version of

11 Assume that Laurence had intended the works to be mimetic rather than spiritual, operable, or some other type of regard that is not historically part of our regard-as-a-work-of-art.
This method says to “regard [artworks] in any way (or ways) in which prior artworks are or were standardly regarded or [regard artworks] in some other way in contrast to and against the background of those ways” (“Defining Art Historically” 233). Thus, instead of limiting the notion of regard-as-a-work-of-art to artworks as they have been properly regarded in the past, it is extended to admit entirely unpredicted types of regard so long as the artist intends to take up regards in conscious opposition to those past correct ways. Consequently, intent for regard-as-a-work-of-art becomes a relational (albeit confrontational) rather than a teleological construction. If the artist at least consciously nods in the direction of past art, her intent for regard is sufficient to grant that the object be regarded as revolutionary art (“Defining Art Historically” 233).

Now that we have established Levinson’s theory, we must test it by considering relevant objections. The first major objection is Steven Davies’s objection, which concerns the historical chain of the ur-arts. He states that “first art must be art already at the time second wave [that is, non-ur-art] pieces become art, otherwise second wave pieces could not be art as the result of standing to first art in the art-defining relation” (Davies, “First Art and Art’s Definition” 21). According to Davies, Levinson’s theory offers no explanations as to how the ur-arts came about, what they are, or how a recursive definition can ultimately depend upon such a vague and undefined idea. Levinson does offer a solution: ur-art procures its art status ex post facto, subsequent to founding of the art tradition from which it stemmed (“Art Historically Defined: Reply to Oppy” 157). However, as Davies points out, this conception of the ur-arts conflicts with a recursive definition, as it is only retrospectively, following the historical chain of regard-as-a-work-of-art, that one may deductively identify the ur-arts. It is not a matter of their retroactively becoming those ur-arts (“First Art and Art’s Definition” 24). Thus, not only do we need an explanation of the ur-arts, we also need one that is recursive.

---

12 It’s unlikely that Levinson believes that this strategy is essential since he denies that revolutionary artists have this intent: “The art-making intention of consciously revolutionary artists may thus have to be a covertly disingenuous one, somewhat along these lines: my object is for regarding in any way artworks have been regarded in the past (but with the expectation that this will prove frustrating or unrewarding, thus prodding the spectator to adopt some other point of view—this being my ultimate intention)” (“Defining Art Historically” 223).

13 This work and Davies’s *Definitions of Art* will be cited by title followed by page number.
Although Davies’s objection may seem convincing, it is refutable. Although the ur-arts do require some type of intent for them to be ur-arts, they do not necessarily need to be regarded in the same way in which subsequent art is regarded. This is because there was no art that preceded them (“The Irreducible Historicity of the Concept of Art” 370–71). The intent that led to the ur-arts was original, as were the ur-arts themselves, but the ur-arts were not artworks in the proper sense because they could not be regarded in the same way that preceding artworks were, because there were no preceding artworks. Thus, the temporal projection takes place as follows: intent for regard in a specific way, ur-arts, intent for regard in the way ur-arts are regarded, first artwork, and intent for regard as previous artworks were correctly regarded, artwork, and so on.

But attacks on Levinson’s theory do not stop there. Victor Haines has raised two important counterexamples to Levinson’s intrinsic recursive definition, one concerning quasi-dollars and the other concerning quasi-vessels:

1. Dollar bills printed by the mint to be regarded as currency that cannot be counterfeited, for example, might be art because it is discovered in some ancient culture that the medium of exchange was art correctly and integrally regarded for its qualities as such a medium (Haines, “Refining Not Defining Art Historically,” 237).14

2. Similarly, bottles and pots intended for regard as functional vessels to hold stuff may be art because in some previous culture containers were art correctly and integrally regarded for the way they hold stuff (Haines 238).

Both of these supposed counterexamples, though, mistakenly suppose that quasi-dollars and quasi-vessels are correctly regarded as works of art (“A Further Fire: Reply to Haines” 77). For example, if a Chevrolet El Camino were simultaneously regarded as a truck and as a car,15 then there must be some other regard (a truck bed) that warrants its status as a truck, aside from a car body (which guaranteed only its status as a car). In other words, if quasi-dollars were art, there must be some reason why they are regarded as art (rather than as currency that could not be counterfeited).

---

14 This work will be referred to by author’s last name and page number.
15 This is a car combined with a truck. It has the body of a car and the bed of a truck.
which goes beyond their regard as currency. Haines has not shown how this is possible ("A Further Fire: Reply to Haines" 77). Although anything has the potential of becoming art, it is not very likely considering the strict recursive criterion, namely a regressive chain from current regard-as-a-work-of-art to one or more \textit{ur}-art, which the object must meet. Moreover, if the culture that regarded quasi-dollars and quasi-vessels as art were purely alien, so that it would not be involved in our current regard-as-a-work-of-art in the evolutionary sense, then there is no reason to regard quasi-dollars and quasi-vessels as artworks within our culture ("A Further Fire: Reply to Haines" 77). Although Levinson’s definition may be recursively applied to other cultures as well, it is “our concept—the Western, Renaissance-derived notion whose current state of evolution [he] was trying to analyze” ("Extending Art Historically" 413).

Other objections to Levinson involve our ability to identify artworks. According to Robert Stecker, we may never be certain whether particular objects, especially those far abstracted from us in time or place, are truly works of art (Stecker, “The Boundaries of Art” 269). Because it is almost impossible to obtain and follow a chain of causality, as one could do with questionable present-day artworks, it would be impossible to envision the intent for regard with which these far-removed objects were created. So Levinson’s intentionalist-historicist account must clearly be wrong.

This objection is persuasive because, as is noted by many philosophers of history,\footnote{See W. H. Dray, Carl Hempel, Michael Oakeshott, and Michael Stafford.} we may never know for certain whether an event of the past was intentional, undesigned, religious, and so forth—intent is irrecoverable. Nevertheless, do we stop trying to deduce historical causation because of the difficulties thereof? Certainly not! Why should the case be any different with artistic intent? Moreover, there can be no doubt as to whether works such as the \textit{Odyssey} were produced with at least a generous number of intentions that we now view as paradigmatically art-making. Most could assume that those intentions were grounded in, and implicitly referential to, those of some earlier enterprise ("Extending Art Historically" 414). Moreover,

\footnote{This is because, just as in geological evolution, the development from the \textit{ur}-arts to our present-day regard-as-a-work-of-art, if isolated, plausibly developed completely unique types of regard. Essentially Haines is asking the wrong question. Rather than giving us a way to incorporate a newly discovered culture’s art into our own definition, Levinson is giving us a way to define our own art (as well as others) historically, given that it is restricted to their specific recursive account.}
given its axiomatic qualities such as exquisite form, extended imagery, and detailed characterizations, there is hardly a way in which we could mistake the work for some other intention. Similarly, we know that the Odyssey must have been embedded in a tradition of anterior endeavors that were regulated by certain intentions that we correctly label artistic ("Extending Art Historically" 414). Even if Homer’s Odyssey were to be regarded as a religious object or a mere chronicle of events, the way that it is written suggests not a single, isolated regard, but a complete ensemble of regards, an ensemble that would include, among its many regards, regard-as-a-work-of-art ("The Irreducible Historicality of the Concept of Art" 370; "Refining Art Historically" 24).

Stephen Davies raises another objection against Levinson’s intentionalist-historicist account:

It becomes difficult to see how Levinson can distinguish the following cases: (a) that in which an artist creates an artwork in the manner in which Levinson suggests; (b) that in which an artist intends something to be viewed as a work of art (and thereby to make it one), when in fact, unbeknown to the artist, it is one already (Definitions of Art 171).

So Davies raises the point of a conflict between a constitutive intention and the nettled issue of proprietary-right conditions. If an object is already an artwork in the correct historical sense, can another artist transform it into another distinct work of art? Moreover, can he do so if he does it unknowingly but nevertheless without a proprietary right over the object?

In order to answer this question, one must first consider the ontology of the art-making involved. Is the artwork simply being used in a conceptual performance through being incorporated, unharmed, into another artwork, or is it being altered beyond the replacement of the original artifact? If the former then, because the second order artwork is simply being erected upon the framework of another artwork without canceling the first order work’s status, there seems to be no problem of a proprietary right.

---

18 Note here that it would not matter if Homer’s intent was to be artistic at that time, as these intents for regard are ones that we currently espouse as historically defined artistic regards (See art-unconscious intention above).
19 See, for example, *Erased de Kooning Drawing*, 1953, in Battin, Fisher, Moore, and Silvers’s *Puzzles about Art: An Aesthetics Casebook* 24 (hereafter referred to by authors’ names and page number).
(“Extending Art Historically” 416). However, if the latter, the artist does have a proprietary-right problem. A commandeered artwork which is physically altered, even if negligently obtained, is still a violation of a proprietary right, and therefore cannot be deemed an artwork (“Extending Art Historically” 417). For example, if I were to dance the Macarena in my bedroom but then add a few elbow touches and knee slaps, the artwork, the Macarena itself, would not be damaged—the art-making therefore goes on unimpeded. However, if I were to expropriate the Macarena, to change the Macarena itself to include my elbow touches and knee slaps into the dance, as it is understood within the fine art tradition broadly construed, the Macarena itself would be damaged, and my alteration is therefore detrimental to its creators’ intentions. No matter what the improvement, if a proprietary right is violated, the offending work is necessarily not a work of art (“Defining Art Historically” 228).

Davies offers another, final counterexample to Levinson’s account, a counterexample worthy of our attention. It is as follows:

There is nothing to stop an aesthetically minded tour guide showing to the members of her tour party some natural scene, the Grand Canyon say, and intending that they regard it in a way that is the correct manner for regarding some type of art, such as an earthwork sculpture (in an impressionist style). The guide’s intention is no wild hope, because the scene invites just that such regard. We would not be that inclined to say the aesthetically minded tour guide turns the Grand Canyon into an artwork (Definitions of Art 173).

But this counterexample can be refuted in a number of ways. First, the tour guide’s intention is merely a passing one, thereby negating the art-regarding intention. Secondly, it seems implausible that the tour guide would in fact be intending for regard-as-a-work-of-art as it is correctly understood (“Extending Art Historically” 418). What historical tie does she present, specific art-conscious intention? Surely not! This is because other earthwork sculptures were trying to present earth-altered impressionist intentions. So, is she presenting a non-specific art conscious intention? Again the answer is no. “Intending something for regard as an x (or as x’s are regarded) is not the same as intending for regard as if it were an x” (“Extending Art Historically” 418). Maybe she is presenting
an art-unconscious intention. But, again, this is absurd since her intent is explicitly conscious, not naïve or transparent. Third, of course the tour guide could make the Grand Canyon into a work of art conceptually, intending the presentation of her tour to be a conceptual artwork properly documented. But this is doubtful because of the way in which she referenced the intention (“Extending Art Historically” 418). Last, and most importantly, even if the tour guide were able to fulfill the qualifications mentioned above, there would still be the issue of a proprietary right. As Levinson aptly points out, because the property upon which the tour guide is attempting to confer her intention is of such a scope that no one could possibly possess it, a proprietary right disqualifies her intent (“Extending Art Historically” 418).

In the above paragraphs we have dealt with several major objections to Levinson’s theory. We have shown that Levinson can overcome these objections. In what follows I would like to point out two more counterexamples that challenge Levinson’s intentionalist-historicist account. The first, though refutable, acknowledges a disturbing outcome of the intentionalist-historicist account, as well as an unanswered query. The second counterexample leads to a contradiction, although a reparable one, within Levinson’s definition.

The first of these counterexamples concerns body art, specifically tattooing. One may claim that a proprietary right becomes jeopardized when a tattoo artist tattoos a client. Because a person cannot own another person, would Levinson allege that body art, being intended for regard-as-a-work-of-art in the same way which prior artworks were correctly regarded, does not qualify as artwork unless the artist is tattooing himself? Well, although the two aspects definitely do clash to some extent, by paying for and approving of the artwork to be assigned to a specific area of skin, the client simply transfers his proprietary right to the tattoo artist, just as a commissioned sculptor is provided a block of marble.

As Davies observes, a sculptor commissioned to produce a sculpture does not have the “right to dispose of the marble as one who has a true proprietary right over the marble would” (Definitions of Art 176). But this protest misfires when Levinson distinguishes between the two by stating that the proprietary right included in his definition does not include the “right to dispose of” in the sense of “placing, selling, or exhibiting the
resulting object” (“Extending Art Historically” 418). Rather, Levinson’s definition refers to the “right to dispose of” in the sense of “using, modifying, or working alterations on certain materials” (“Extending Art Historically” 418). Thus, Levinson’s definition seems to include tattooing.

What if, however, a person was sedated and a tattoo artist tattooed a beautiful rendition of Van Gogh’s *Starry Night* upon him? In this case, according to Levinson, the work would not be an artwork because of the proprietary right. At this point, all that I can offer in defense of the tattooed *Starry Night* rendition is that it seems upsetting that such a clearly aesthetically pleasing work will not be included in the historical definition whereas Levinson has made room for inclusion of a urinal on its side (Battin, Fisher, Moore, and Silvers 16). Even so, this example raises yet another alternative state of affairs, one that Levinson has not previously addressed. Suppose, upon waking up, the tattooed victim really liked the piece and thanked the artist for doing such a great job. What if an artist intends for regard-as-a-work-of-art an object that he does not have a proprietary right over and then, upon discovering the work, the person who does have the proprietary right approves of the work? How do we account for the position of this work where there is no explicit nor implicit transfer of a proprietary right?

The second counterexample involves graffiti. Noël Carroll argues that, contrary to Levinson, “had Picasso stolen into a subway yard at night and, after the fashion of graffiti artists, painted *Guernica* on the side of a

---

20 In fact, for our purposes, an original artwork would be just as effective. Here I am simply using the example of a widely known artwork so that there is no confusion about what exactly the tattoo artist is tattooing.

21 It seems to me that what Levinson is referring to as “self-aware, revolutionary arts” are precisely works such as Duchamp’s *Fountain*. Thus, the whole reason for proposing the liberalized version of regard-as-a-work-of-art is aimed at the inclusion of Duchamp and similar artist’s works, whereas the proposed example is excluded. Levinson’s definition obviously does not make room for everything and anything that we think is pretty. However, I personally find it disturbing that works that are aesthetically pleasing are not given special consideration, whereas Levinson makes room for pieces that are designed with the intent to specifically aggravate aesthetic sensibilities. Nonetheless, my subjective objection has no logical grounds; Levinson’s definition is sound (albeit with the subtraction of the liberalized regard-as-a-work-of-art).

22 It has been pointed out that the victim’s appreciation may constitute an implicit, if ex post facto, transfer of proprietary rights. However, I do not believe this is relevant because, necessarily, an ex post facto transfer of the proprietary right took place after the work was completed. The artist’s work cannot be considered art because his intent did not meet the proprietary-right condition. Remember, it is not the work itself that makes it art, but the artist’s intent for regard at the time of conception.
train, it would be art no matter what Mayor Koch says" (Carroll, “Art, Practice, and Narrative” footnote 9). Levinson does not agree: “stylish letterings and designs on subway cars produced by graffiti artists who had no right to deposit them there would automatically not be art by my lights, given a proprietary condition over materials” (“Extending Art Historically” footnote 20). Although Carroll does not pick up his argument in favor of graffiti beyond this context, with respect to Levinson’s definition at least, I would like to do so here. It has been noted by art historian Kevin Element that:

[Graffiti] artists are yelling out, not as immature youths, but rather as mature modern thinkers during a period of history when artistic dissent and creativity is stifled by an art world filled with dogmatic tradition and a media culture pathetically addicted to the consumption of messages and images designed to propagate the sales of consumer products (Element, “Hard Hitting Perspectives on Hip Hop Graffiti”).

According to many who study the graffiti art subculture, graffiti artists’ purpose is not to simply vandalize but rather, as was done by Duchamp and the rest of the Dadaists of the early twentieth century, to deny the applicability of all past ways of looking at art objects. However, whereas the Dadaists took the conceptual route to express their message against the art world, graffiti artists take the material route by denying that materialism (the acquisition of artistic possessions) has any place in the art world. In fact, as stated in the second issue of Xylene, a graffiti zine out of Vancouver, British Columbia, “Art is very powerful and should be for the masses, not the elitist. . . . We must break from the art world and create our own separate identity. . . . Graffiti will be on every surface except your television” (Anonymous, “Major Writers Meeting in New York City in the Year 2000” 18).

There could be no clearer example of an artwork which is intended “to adopt such regards in conscious opposition to those past correct ways [of regarding art]” (“Defining Art Historically” 233). All that Levinson requires in his liberalized version of regard-as-a-work-of-art is that “[the artist] should project the new way [of regard] in . . . antagonistic relation . . . to its predecessors” (“Defining Art Historically” 233). However, this causes a

---

23 See Beaty and Gray’s “Zap! You’ve Been Tagged”; Brewer’s “Hip Hop Graffiti Writers’ Evaluation of Strategies to Control Illegal Graffiti”; and Chalfant’s Spraycan Art.
problem because artworks, as they are correctly regarded within Levinson’s definition, must meet a proprietary condition. Thus, if one were to truly regard-as-a-work-of-art in contrast to the correct regard-as-a-work-of-art, as it is understood by Levinson’s own definition (containing a proprietary condition), the opposing regard-as-a-work-of-art would necessarily not include this proprietary condition. Graffiti, by specifically targeting the proprietary condition and other aspects of the background of correct regard-as-a-work-of-art, qualifies as art under the criterion of the liberal version of regard-as-a-work-of-art. However, this leads to a contradiction because the proprietary condition is a vital component to Levinson’s historical definition; graffiti does not meet this condition.

Although the above graffiti counterexample leads to a contradiction within Levinson’s definition, said definition can be successfully reformulated. But to reformulate it either the proprietary condition or the liberalized version of regard-as-a-work-of-art must be dropped from the contradiction. As I have shown, Levinson’s definition is robust barring the above stated counterexample. Thus, as the most likely candidate for a recursive definition of art, it is important that we address this contradiction. One way to do this is to drop the least necessary side of the conflict.

Since this problem arises because of the proprietary condition, let us first discuss some consequences of the proprietary condition and then see how we can fix the problem. As seen in response to Davies’s counterexample above, a proprietary condition limits what one can regard-as-a-work-of-art to objects that can actually be possessions. Artists cannot simply turn the sky into a work of art, because the sky cannot be owned by anybody. Moreover, as observed in reply to another of Davies’s objections, a proprietary condition limits artworks to those works which are not commandeered and then physically altered. An artist cannot merely walk by another artist’s work while splashing paint and call the product an artwork of his own. Last, a proprietary condition limits curators, promoters, and exhibitors from turning non-art objects of the past into art.

What, though, are the consequences of the liberalized version of regard-as-a-work-of-art? It seems that the only role for the liberalized version of regard-as-a-work-of-art, at least the only one that we have observed thus far, is the inclusion of truly revolutionary works of art. It can be argued that if Dadaism was truly revolutionary in that its members were
manufacturing works not regarded in any way other than counter-relationally to regard-as-a-work-of-art in which prior artworks are or were correctly regarded, these works would not be included within Levinson’s definition without the liberalized version of regard-as-a-work-of-art.

Based upon the consequences of both the conditions above, we must conclude that the most profitable component in question is the proprietary condition. This is because, as noted above, Levinson’s liberalized version of regard-as-a-work-of-art is under-limiting. Under this version of the definition, any object with the artist’s intent being that of upsetting the art world is acceptable. However, this version of the definition is, in actuality, rarely necessary. According to Levinson himself, even the most revolutionary artists may still evoke a historical regard without the liberalized version of regard-as-a-work-of-art. For example, an artist’s intent may be that his or her “object is for regarding in any way artworks have been regarded in the past (but with the expectation that this will prove frustrating or unrewarding, thus prodding the spectator to adopt some other point of view—this being my ultimate intention)” (“Defining Art Historically” 223). Even without the liberalized regard-as-a-work-of-art, most revolutionary artworks may still be included under Levinson’s definition, but, in order to defeat the graffiti counterexample above, we must drop the liberalized version of regard-as-a-work-of-art from a historical definition.24 However, I do not believe that this expulsion will substantially effect the historical definition because, as explicated above, Levinson still has one scheme available that may allow for the inclusion of revolutionary artworks: such works are art if they initially direct their audience in a way in which artworks have previously been regarded (“Defining Art Historically” 233).25

In sum, I have addressed the more substantive objections and counterexamples to Levinson’s definition and came out with a distinguished

24 It may be noted here that by retaining a proprietary condition graffiti does not qualify as art. Although some, for instance Carroll, may still object to this conclusion, the point of the counterexample was to expose a contradiction between two components of the historical definition, not to extend the definition to include graffiti. According to one aspect of Levinson’s definition graffiti is art. However, according to another aspect (the proprietary condition) graffiti is not to be considered art. The graffiti example is not meant to contradict Levinson’s entire theory, but rather to expose a contradiction within the definition so that we may address it.

25 For example, an artist may sign the artwork in the lower right hand side, as per custom within the art world, or he may, as the Dadaist artists did, name the work in a way that has been historically associated with artworks. This would include Dadaist artworks, but not, because we have dropped the liberalized version of regard-as-a-work-of-art, graffiti.
and reinforced set of limits upon which to rest a definition of art. However, by utilizing graffiti as a counterexample I exposed a contradiction between Levinson’s liberalized version of regard-as-a-work-of-art and his proprietary condition. All the same, this contradiction did not prove devastating to the historical definition because we were able to drop the liberalized regard-as-a-work-of-art from our historical definition without necessarily harming the definition at all. The expulsion does, however, clean up an already formidable anti-anti-essentialist definition of art. As stated by Levinson himself, “Which of the two [revolutionary] strategies is ultimately preferable as a way of accommodating the historical definition of art to revolutionary art making is a question I will not settle here” (“Defining Art Historically” 233). I believe that we have settled that question here.
Works Cited


Hacking and Human Kinds

ENOCH LAMBERT

Ilan Hacking says that human kinds are different from natural kinds.¹ What kind of a claim is this? What import is it meant to have? Is it a metaphysically innocent claim comparable to stating that biological kinds are different from chemical kinds, which are different from subatomic kinds (assuming, for our purposes, the metaphysical innocence of these claims)? Is it a metaphysically forceful claim—one stating that human kinds differ from natural kinds in some essential respect? Or, is it somewhere in between—one implying that, though there may not be any “deep” metaphysical difference between human and natural kinds (e.g., that humans are somehow radically non-natural), theoretical inquiry into human kinds must nevertheless function in some strikingly different ways and with different results from all other kinds of natural theoretical inquiry? Hacking has not been very clear on this point. Though he tries to avoid making statements explicitly committing him to strong metaphysical positions, his arguments do seem to carry metaphysical implications. On a metaphysical commitment scale of one to five, with one being the innocent claim and five being the forceful claim, Hacking’s writings on human kinds

¹ See especially “Making Up People,” “A Tradition of Natural Kinds,” “The Looping Effects of Human Kinds,” and Historical Ontology. These works, as well as the rest of Hacking’s works referred to in this paper, will be cited only by their titles. Few page numbers will be given because I am generally citing the main idea of the relevant texts rather than specific passages.
strike me as placing him somewhere around a four (I use this scaling technique in imitation of and response to Hacking’s suggestion in *The Social Construction of What?*, where he uses a similar one himself in regards to social constructivism).

Hacking’s writings on human kinds rely very much on observation and empirical methods. Quite often, he seems not only to be documenting the interesting differences in human kinds he claims to have found, but to be inviting or even urging others to pursue his particular kind of research (his writings are often peppered with new hypotheses and provocative questions).² In one of his most important pieces on this topic, Hacking even calls his work a contribution to the “study of making up people” (“The Looping Effects of Human Kinds”). Tendencies such as these place him squarely in the three range of the metaphysical commitment scale described above. However, if the results of his inquiries are true (as I think they are), there is no denying that they imply things about human beings that do not sit well with certain tendencies in the naturalistic tradition in philosophy. Insofar as this is so, Hacking’s position gets nudged into the four range of the scale. In this paper I would like to further develop and defend some of the lines of reasoning that lead Hacking to the positions he takes. Specifically, I will defend the thesis that human kinds differ from other natural kinds in an ontologically significant sense by arguing that the epistemology, semantics, and ethics of human kinds feed back onto and change their ontology in a way unlike other natural kinds. Developing and extending Hacking’s use of the existentialist tradition in his account of human kinds, I will argue that existential meaning is one of the primary mediating factors in this kind of feedback. Using existentialist insights into identity and action, I will also more fully develop Hacking’s use of the idea of spaces of possibilities³ for action in a way that makes the idea less reliant on Anscombe’s work on intentional action and so less open to criticism due to Anscombe exegesis.

² One might even wonder if Hacking is trying to develop what he calls a new “style of reasoning” (*Historical Ontology* 2).
³ This is my gloss on Hacking’s ideas, an attempt to bring them under one category.
Making Up People Through Looping Effects

Hacking says that we can and do “make up people” or human kinds (“Making Up People”). By this he means that we make up new identities and personalities as well as new types and styles of actions, emotions, and experiences to go along with them. To illustrate this point, he considers Sartre’s famous narration of how a French garçon de café attempts to fulfill the role of being a French garçon de café. Hacking points out how “being a French garçon de café” could not have been anywhere on the map of possibilities for a person in France just a few hundred years ago. Likewise, even though we know that there were French lords and vassals, being a French (or any other) lord or vassal is not anywhere within the range of possibilities for anyone today. These are the types of human kinds that Hacking thinks get made up (as well as disappear). Examples of human kinds that Hacking thinks have been made up relatively recently (i.e., at various times within the last 175 years) and for which he has done historical and empirical research include people with multiple personality disorders, homosexuals, child abusers, schizophrenics, and autistic children. While he does not say that these kinds of people are not real—schizophrenia and child abuse, for example, can be all too real for those who suffer from them—Hacking does give evidence for each of these kinds being relatively new phenomena as kinds (i.e., classifications and identities toward which people meaningfully relate themselves) and asks how it is that they came about. Note that Hacking’s interest here is in some ways rather limited. He is not talking about all human social phenomena. He is not talking about, for example, artifacts, languages, nations, or even many kinds of people. He does not talk about parents, leaders, or farmers or other nearly universal human kinds that exist, in some form or another, in most cultures and in most times.

How does Hacking think human kinds of the sort mentioned above get made up? He says that they arise through a distinctive kind of “looping effect” that obtains between, to use his words, “culture and cognition,” or, to put it another way, between everyday practices and classificatory practices

---

4 Hacking explicitly prefers this phrase to “socially construct people.”
(see “The Looping Effects of Human Kinds”). Once a human kind gets classified and becomes an object for knowledge, control, healing, punishment and so forth, the kind will tend to react and respond to such treatment. The reaction thus creates new properties of the kind. In this way a looping effect is initiated in which kinds and knowledge create each other. Some commentators such as Cooper have focused almost exclusively on classification by theoretical disciplines in the feedback system, but I think it is important to note that Hacking often emphasizes governmental classification. For instance, we learn that obsessive, nineteenth-century governmental classification of suicide motives was part of a feedback loop that helped create many of the phenomena surrounding suicide (such as the suicide note) that were already in place when Durkheim performed his famous studies of it (“Making Up People”). However, while it is very important to keep in mind, the full significance of this distinction between kinds of classificatory practices in the production of feedback loops will not be my primary consideration.

Hacking’s examples of the effects of feedback loops are surprising and stark. For example, he cites evidence to the effect that the current concept of child abuse did not come into being until about 1961, as earlier concepts of cruelty to children did not include sexual abuse (see Rewriting the Soul). Soon afterwards, when governmental agencies started keeping records, the number of child abuse incidences reported was in the low thousands. In less than twenty years that figure had ballooned to over one million and to over two million less than ten years after that. Hacking reports wildly different estimates of what percentage of children have actually been abused. He talks about how child abuse became just about the worst evil one can perform, how now children learn about it in schools, and so on. Through historical investigation, Hacking shows how what is now a very “relevant kind” to our culture was, until very recently, not a part of it at all. The historical chain certainly makes it seem as though a new classification or representation kick started a significant cultural phenomenon.

5 See Cooper’s “Why Hacking is Wrong about Human Kinds” (hereafter referred to only by author’s last name).
Realism and Idealism about Kinds

The contention that classification and representation affect their objects immediately arouses the specter of idealism. Is Hacking’s proposal about human kinds another form of Kantian idealism? Does the representation of human kinds constitute them? Boyd has offered a distinction that will be helpful here (Boyd, “Realism, Anti-Foundationalism, and the Enthusiasm for Natural Kinds”). In a paper responding to one in which Hacking talked about the looping effects of human kinds, Boyd distinguishes between social practices (such as legal classification) and theoretical disciplines (such as sociology) causally affecting and interacting with their objects of study and non-causally doing so. He says that only positions holding that social practices and theoretical disciplines non-causally affect or constitute their objects (through convention, analytic truths, etc.) should be considered idealist or neo-Kantian. Boyd offers Kuhn’s position of paradigms constituting new worlds as an example. On the other hand, he thinks that any position contending only that there is some causal interaction between theoretical or classificatory practices and their objects of study is perfectly consistent with a realist position. Hacking’s later article on looping effects stresses the fact that Hacking thinks the looping relationships in human kinds between culture and cognition are causal (he even stresses that the relevant causes are efficient causes) and that the goal in studying them is general knowledge (“The Looping Effects of Human Kinds”).

Of course, there have to be some restrictions on the causal interaction of things and their theoretical study and classification for realism to be a viable position. One cannot hold that theoretical disciplines causally change whatever is essential to their objects of study and still hold that such disciplines only discover (and not also create) truths about their objects. Boyd formulates this restriction through his “accommodation thesis” (see, for example, “Homeostasis, Species, and Higher Taxa”). In order to be successful, he says, our explanatory and inductive practices must accommodate themselves to the causal structures of the world. The use of our terms that refer to things in the world must be, at least in part, causally

---

6 This paper and others by Boyd will be cited only by their titles.
regulated by those things themselves and their causal structures. When we accommodate our explanatory and inductive practices to these causal structures (whatever they may turn out to be), there is no need to worry about those same practices causally affecting them to any significant degree. The question, of course, becomes whether such accommodation is possible in the study of Hacking’s human kinds.

Another way of putting the realist problem is to say that, for the natural sciences, the epistemology, semantics, and ethics of natural kinds do not fundamentally affect or condition their ontology. However, as I said above, it is precisely my position that Hacking claims that the epistemology, semantics, and ethics of human kinds do affect their ontology. How, then, can he be a realist? Two papers critical of Hacking, Boyd’s “Realism, Anti-Foundationalism, and Enthusiasm for Natural Kinds” and Cooper’s “Why Hacking is Wrong about Human Kinds,” raise the question of realism for Hacking. Boyd does not think that Hacking’s considerations should keep us from extending “enthusiastic realism” from natural to human kinds. And Cooper thinks that Hacking’s position, if true, would have negative implications for the possibility of knowledge, induction, and the formulation of laws in the social sciences (Cooper). Part of the reason for their concern, I think, has to do with ignoring the limited nature of Hacking’s claims that I mentioned above. Boyd speaks of the “possibility of social knowledge” and Cooper speaks of the “social sciences” in general. But, as I have pointed out, Hacking is concerned about only some human kinds—not all human kinds and certainly not all human social phenomena. They also seem to ignore the fact that Hacking has actually offered quite a bit of general knowledge about the very examples of human kinds he talks about. He has done historical research documenting the rise of each of the kinds he discusses as well as their general features. What is more, he claims to have some general knowledge about their causal structures—that they are constituted through looping effects between “culture and cognition,” that part of the causal mechanism of these looping effects includes modification

---

7 The editors, in commenting on this paper, raised the important issue of whether all human kinds are subject to Hacking’s looping effects. Here I am focused on Hacking’s existence claim that at least some human kinds are subject to looping effects. The question of if or how one could distinguish between human kinds which are or are not subject to looping effects is an important question but not the topic of the current argument.
of possibilities for being and action, etc. Thus, the worry about radical skepticism and strong idealism about human kinds should be defused.

Nevertheless, I do think Hacking is committed to holding that there are important limitations to our knowledge of human kinds as well as differences in how we arrive at that knowledge. First, there probably are differences between the study of “made up” human kinds and other natural kinds. The influence of Foucault on Hacking’s project (which he explicitly recognizes in *Historical Ontology*) should be noted here. Foucault certainly did not think that his “archaeologies” and “genealogies” of human knowledge and kinds were to be seen as modeled on the natural sciences, though he did think he was uncovering real knowledge nevertheless. Second, there probably are no laws associated with the making up of human kinds. Even when a causal mechanism has been identified (e.g., the looping effects), this does not mean that there are strict laws for it. For example, there are no universal laws like the laws of physics that could be used to predict and explain each and every emergence of a new human kind. As Hacking himself says, a unique story probably needs to be told about each human kind (“Making Up People”). For human kinds, then, narration and interpretation are probably ineliminable aspects of appropriate explanation. Third, and most important for this paper, human kinds may be so affected by classificatory practices and theoretical inquiry modeled on the basis of the natural sciences that any claims about the essences of such kinds based on such inquiry is likely to be more creation than discovery. For example, whereas chemistry may discover that water is necessarily $\text{H}_2\text{O}$, no such claims about the fundamental identity or causal structure of human kinds such as schizophrenia may be forthcoming. It is to this possibility that I turn next.

**Existence, Essence, and Epistemology**

As I have mentioned, Hacking employs some of Sartre’s thoughts to illustrate his point about human kinds. Elsewhere, he explicitly recognizes existentialist influence on this work (see *Historical Ontology* 22–23). This is very important in understanding what Hacking says about action and being. One of the core principles of existentialism is that, for humans,
existence precedes essence. Hacking interprets this principle as meaning “we are constituted by what we do” (Historical Ontology 22–23). Of course, for Sartre, “what we do” is completely up to the subject because Sartre thinks the subject is completely free to impose whatever meaning it wants on what it does. Hacking, on the other hand, stresses that our possibilities for action are socially structured and constrained. This view is a more Heideggerian way of looking at the problem.\(^8\) And there is more in what Heidegger says that I think is relevant to Hacking’s project.

As opposed to Sartre, who thinks that the “existence principle” precludes humans having any essence, Heidegger argues that a human’s (Dasein’s) essence is to be found in his existence. For early Heidegger at least, this means that part of the essence of what humans are lies in their being humans (Heidegger, Being and Time).\(^9\) This means, I propose, that while no description nor any theoretical identity statement (like “water is H\(_2\)O”), can ever fully capture or identify the essence of being human, humans still share something of a common essence in virtue of their being human. The only way to know fully the essence of being human, though, is to be one. Call this idea the Non-Representability Thesis (NRT). This thesis, then, implies quite the opposite of saying that representations constitute their objects. Rather, the NRT claims that there is something to being human that is in some sense real that cannot be theoretically represented (though Heidegger argues that what is real in this case is not an entity but a way or mode of being). Obviously, it is important to be clearer about the nature of this claim. For example, the NRT is not meant primarily as a kind of mystical ineffability claim to the effect that human lived experience is kind of like a representable experience, only “souped-up” enough to be just out of reach of normal human representation. Rather, it is meant to claim that the existential meaning of being human is non-transferable: the essence of the meaning of being human and the particular way it motivates me is my responsibility and cannot be given to any one else. Descriptions and representations are tools for transferring and disseminating things like bits of information. What cannot be

\(^8\) This is not to say that Hacking is influenced by Heidegger. Indeed, Glazebrook says that Hacking told her he had decided never to read Heidegger. On the other hand, as Foucault was very much influenced by Heidegger, there may be some indirect influence (Glazebrook, “Heidegger and Scientific Realism”).

\(^9\) Hereafter simply “Heidegger.”
transferred, however, is my personal responsibility for my particular stand on what it is to be a human being. To give a more concrete example, I can complain all I want to my friends and relatives about how the stress of finals is getting to me, and they can all empathize with me (the NRT is not a solipsistic doctrine). In the end, however, only I can take the responsibility for the choice of how to deal with the stress by studying hard, taking the finals, not taking them, etc. And only I cannot escape the way my choices and their consequences will help condition my life and the meaning it has for me. The NRT rests on the claim that there is something essential to human being about the non-transferability of responsibility for one’s choices and way of being. Note that most existentialist writers (including such figures as Kierkegaard, Nietzsche, Heidegger, Sartre, and Camus) have written in a wide variety of styles, using new vocabularies, aphorisms, plays, novels, and pseudonymous authorship in order to try and engage the implications of the NRT and to avoid things like theoretical identity statements for human beings.

I think that one way of looking at what Hacking is doing (or is committed to even if he would not say so), in effect, is extending the NRT from humans in general to particular human kinds. Part of the essence of being homosexual or schizophrenic consists in the ways in which the meaning of being those things conditions a person’s life, her motivations, possibilities for achievement, fulfillment, action, and so on. Note that the NRT entails that there is some aspect of human kinds unavailable to theoretical knowledge. Furthermore, because there is some such reality (being human) that is not fully representable but that does in fact causally (or motivationally) interact with or is part of the feedback loop, strict laws about how feedback loops work are probably impossible. Let us call that essential aspect of human being that is not fully representable—I have not said anything implying that it cannot be talked about at all—“existential meaning.” In the ensuing discussion I will sometimes use the word “motive” in suggesting how existential meaning operates in people’s lives. By “motive” and its cognates I mean those meaningful possibilities for the sake of which an agent can act.10 In order to defend Hacking’s important points in this regard, I would

10 Though this way of considering motives is not unique to him, I primarily follow Wrathall’s “Motives, Reasons, and Causes” here.
like to turn to how existential meaning mediates the conditioning of possibilities of action.

Meaning, Possibility, and Action

Hacking employs Anscombe’s work on intentional action to show how concepts and meaning can have causal efficacy in human kind feedback loops (“Making Up People” and “The Looping Effects of Human Kinds”). Because, according to Anscombe, all action is action under a description, new descriptions mean new actions. And, because the “existence principle” indicates that we constitute ourselves through what we do, new actions mean new identities. New identities can be classified and so new identities mean new kinds. Hence, from new descriptions spring new kinds. Considering Hacking’s account of human kinds as a whole, this claim certainly raises the most potential for worries about idealism. Part of what Hacking needs to avoid the idealist charge is an account of how the relevant descriptions and classifications are not arbitrary (so that not just any classification would be causally efficient), even if they are not fully representative of a pre-existing reality. He must also give an account of how some descriptions and classifications take hold while others do not (call this their “stickiness” factor). Later, I will try to give an outline of how an account of the non-arbitrariness and stickiness factors of human-kind-creating descriptions would go. For now, let me try to deflect some other criticisms of Hacking—the charges that he misinterprets Anscombe.

Hacking’s use of Anscombe has been the subject of quite a bit of criticism, and the critics have tended to dismiss his theses about the role of possibilities for action in looping effects because of it (see “Why Hacking is Wrong about Human Kinds”; Sharrock and Leudar, “Indeterminacy in the Past”; and Allen, “The Soul of Knowledge”). I think the point about Anscombe exegesis could be put aside altogether if Hacking used the more existentialist account of meaning and motives outlined above for his theory.

---

11 The editors of this paper wondered if this amounts to a denial of the NRT. In reply I’d like to say that meanings and representations having some causal impact in the constitution of human kinds does not preclude the necessity of the NRT for an account of how human kinds can be constituted through looping effects. Both are necessary but singly insufficient for the final account.

12 Sharrock and Leudar’s work will be referred to by title alone.
of how action and human kinds interact. Let me introduce that account, however, by way of brief consideration of one of Hacking’s critics.

Cooper interprets Hacking as arguing that descriptions extend the logical space of possibilities for action and of using Anscombe’s work to back him up (Cooper). She then criticizes Hacking’s interpretation of Anscombe and argues that he has not made the case that the logical possibilities of action for an agent are dependent on descriptions available to her. She admits that a person’s possibilities for action may be contingently dependent on certain concepts, descriptions, traditions, etc., but thinks this is not metaphysically significant. On the model of logical and contingent possibility, she may be right. But there is a third kind of possibility, existential possibility, for which those categories are not adequate. On the model of existential possibility, what are thought of as merely “contingent” possibilities from a purely detached, theoretical standpoint, may be absolutely essential to the identity of human kinds. And, contra Cooper, Hacking recognizes this and uses it in his account (though he does not use the phrase I do). He speaks of “relevant” and “live” possibilities for action and personhood and even employs some of Sartre’s more dramatic idioms of how certain possibilities for action are “absolute, unthinkable, and undecipherable nothingness” for people in different times and places (“Making Up People”; see also “The Looping Effects of Human Kinds”). However, here again, I think that Hacking’s account can be supplemented more adequately by Heideggerian insights than Sartrean ones.

Existential possibilities for action differ from logical possibilities in the sense that some “show up” as making sense to do and others are not even, so to speak, “on the map” (see Dreyfus’s Being-in-the-World). Some existential possibilities for action cohere with others into spaces of possibilities that make sense for people of certain types to do. Mere logical possibility does not do this. For example, Samurai honor suicides are not within the space of possibilities to a contemporary American. These possibilities also differ from “contingent” possibilities in the sense that they can be life-defining. For example, according to Kierkegaard, for someone who has truly made a life-defining commitment in marriage to another, it makes no sense to say to them “you could have fallen in love with someone else”

---

13 This and other works by Dreyfus will be cited by title alone.
The main point (even if he is wrong about that particular example) is that our actions, though “contingent” in the classical philosophical sense, still thoroughly condition our identities by opening up or closing down further possibilities. For example, though a person might be very talented at multiple sports, the pursuit of becoming a professional at one most likely closes off the possibility of maximizing one’s talents in another. However, these possibilities are dependent on another kind of existential possibility having to do with identity. In the case of both logical and contingent possibilities, possibilities are defined in terms of non-actual, or not-yet-actual states of affairs (substitute events or processes according to your ontological tastes). But possible existential identities cannot be defined in this way. Existentially defining roles, such as being a mother, being a professor, being American, etc., are not states of affairs that can ever be realized and finished. When someone is a mother, and for as long as she is a mother, being-a-mother is always among the possible ways-to-be that open up certain possibilities of action for her. Being-a-mother is inexhaustible. There is no culmination of events in which “being-a-mother” is once and for all fulfilled. It is constantly (so long as a person chooses it) a possibility toward which a mother may meaningfully comport herself. Even possible ways-to-be that people choose to relinquish still condition people’s future possibilities. For example, being a divorced woman has different meanings and possibilities open to it in our society than just being a single, never married woman. It may be countered in response to this line of reasoning that, of course, “identities” are not possible states of affairs, but that is not something unique to humans—being-a-dog is an identity, for example. The difference, though, rests in existential meaning. Being-a-dog is not a meaningful possibility toward which dogs can comport themselves—a dog does not, for instance, take a stand on its particular way of manifesting "doghood." With this account of existential possibility in place, we can now

---

14 I am using Kierkegaard in my “Heideggerian” approach to Hacking due to Dreyfus’s interpretation of Heidegger as being influenced by Kierkegaard.

15 The following explication of this idea rests on my interpretation of Heidegger.

16 As has been pointed out to me, a mother can continue to pursue her identity as a mother after the death of her child. This is because so much of her life and her life’s activities have been structured around being a mother. Studying the Samurai lifestyle in the absence of the live traditions, practices, and concerns of the Samurai, however, could never make one a Samurai. At best, it could make one a Samurai-obsessed person of modernity.
turn to the semantics of identity and action terms and descriptions in order to bolster Hacking’s own account. In looking at how non-exhaustible existential possibilities and actions that are performed for the sake of them interact with the meaning of the terminology used to describe them, we can better understand the “micro-processes” of Hacking’s account of looping effects.

The Semantics of Human Kind Terms

I think that Hacking was on to something when he tried to employ action-creating descriptions in his account of the process through which looping effects make up people. Real, concrete descriptions and classifications are good candidates for the mediation of the effects of abstract things like concepts and meaning in feedback loops. They can also do the job of keeping the relevant action meaningful and not behavioristic. But Hacking needs to make some important distinctions in order to save fruitful debate about his account from descending into endless exegesis of another author, in this case Anscombe. First, there needs to be a more explicit distinction made between terms and descriptions. Second, there needs to be a distinction made, of the kind just discussed, between types of identities (Hacking’s human kinds) and the space of actions possible for them. Finally, we will need to see how these distinctions interact and what they mean for Hacking’s account. An account of how the semantics of human and action kind terms and descriptions work will also help us see how existential meaning is involved in the mediation of the looping relations between classification and culture. For purposes of clarity and presentation of the distinctions, I have included the table on the following page. Some examples of each term kind are taken from Hacking’s own examples of human kinds for purposes of illustration of subsequent points.

In discussing Anscombe on action, Hacking seems only to indicate and discuss the role of descriptions of actions (#3a and #3b of table). But a full account of the causal mechanisms involved in the mediation of the feedback loops between cognition and culture needs to make all four of the distinctions listed above. In the making up of new kinds of people, the creation of new types of all four is potentially very important. Let me discuss them.
First, both actions and kinds can have general terms (#1 and #2). General terms for actions can take the form of gerunds or more regular noun forms. Examples of the former include “killing” and of the latter “speech.” I want to argue that because Hacking discusses only descriptions he leaves out an important property of general action and human kind or identity terms, a property that is crucial to an account of the looping effects of human kinds. At least some terms picking out general types of actions and human identity kinds have, above and beyond their purely referential function, non-referential meaning. That is, to use Millian terms, they have connotative properties above and beyond their denotative ones. This may be in contrast to Kripke’s influential arguments to the effect that proper names and natural kind terms refer but have no sense (Kripke, Naming and Necessity). Whatever the merits of those particular claims are, I think that they cannot be extended to at least some general human and action terms. However, the arguments for this have to do with empirical observations and not just conceptual analysis or consultation of linguistic intuitions.

As Hacking reminds us in his discussions of human kinds, it matters to people whether they are classified by certain terms or not. It also matters

---

17 However, I am not claiming that, for example, the non-referential meaning is determined by something like Fregean Sinn. As we will see, I think that something more like existential meaning is often involved.
to them how their actions are identified. What people are called can matter to their physical well-being (e.g., being called a Jew in Germany in the late 1930s), their emotional well-being, the space of possible actions and other identities available to them, etc. The same goes for their actions. It makes a difference whether a person’s actions are labeled “murder” or “self-defense.” It makes a difference to a homosexual whether his sexual activity is identified as sodomy or love-making. General terms classifying human actions and identities are not purely referential because they have meanings or connotations for the sake of which people can act in the process of working out their own ways of manifesting what it is to be the kind involved. When I say that terms have meaning for the sake of which people can act, I do not mean to imply only favorable or even largely free action. Often, the meaning of terms play a restrictive role in people’s lives, serving as a basis against which people react—and try, among other things, to escape, fight against, or overcome.

Another reason to think that general terms have sense is that the same term, with the same referential function, can have different connotations to different people, at different times, and in different contexts. The term “homosexual” may have very different connotations when used among a fundamentalist Christian sect from when it is used at a gay bar, for example. However, I do not think that the meaning of kind terms is exhausted or fully determined by a set or cluster of descriptions. There are at least two reasons for this. Connotation can be invoked and manipulated by tone of voice, gesture, social setting and so on—that is, by action. Second, their non-referential meaning is often up for contestation in society. Because of this, I will hypothesize that a human kind term’s possible meanings are part of its connotation, where “possible” is meant in a weak sense, covering only those meanings within range of the current debate over the proper meaning. That a term has become politicized and a locus of struggle over identity, control, etc., and so has implications for the future, is part of its connotative meaning. This is another reason why descriptions cannot fully capture the meaning of a term—the cluster of possible descriptions is being contested. This is analogous, I think, to the way identities and actions are related to each other. Identities, as shown above, are constantly available for motivating new actions. No set of actions can exhaust
or complete a kind of identity. Conversely, identities are constantly being affected by actions, descriptions, and terms. The four are in a mutually conditioning relationship. What kinds of actions are essential to an identity? What must a person think or do (if anything) to be a homosexual? What defines a child abuser? How can we know when child abuse has occurred? So often, questions like these become ones of political and social control rather than anything like scientific discovery. How someone is classified will always be a matter of existential import and meaning to them. Let me now briefly say some more about descriptions.

Saying that all action is action under a description, as Hacking does, also obscures two ways in which this might be the case (see #3a and #3b in the table). Action descriptions may either (1) further specify or describe conditions under which an action term is applicable (e.g., “using someone's guilt to make them do something is manipulation”) or (2) report particular instances of actions performed by particular agents (e.g., “Professor Plum committed the murder in the dining room with the candlestick.”) Descriptions of the first kind help determine the range over which general action terms are applied. Descriptions of both kinds often include evaluative modifiers (both adjectives like “intense” or “skillful” and adverbs like “very”). Descriptions of the second kind, as well as their corresponding actions, can also contribute to the general style and meaning of the general type of action through creativity, audaciousness, norm violation, etc. The list of possible properties of action descriptions could go on. The important point to make is that there are probably feedback loops going on all the time at the “micro-level” of descriptions, terms, and actions. Even descriptions of type #3a and #3b and their corresponding actions can influence each other. To say this is to characterize the loop like this: specific actions will have effects on how they are described. How specific actions are described will have an effect on how general types of actions are described and categorized. The categories and descriptions of general action types will impact the meanings of terms for both actions and human kinds. Finally, the meanings of action and human kind terms will have an impact on the way people respond to being classified according to those human kinds (which are themselves specific actions starting the looping process over again). Let us now consider a concrete example.
With these accounts of both existential possibility and human kind semantics in mind, we can look at an example of how classification and action mutually influence each other. We can see how certain terms and possibilities stick and others do not (this is the account of non-arbitrariness and stickiness of terms and descriptions promised above). In discussing Heidegger’s later views on language, Dreyfus gives the example of the phrase “laid back” (“Husserl, Heidegger, and Modern Existentialism”). He says that before that phrase was invented and came into use to describe a style or way of living, people were already engaging in the sorts of practices to which the phrase applied.18 What the phrase did was “gather” the sorts of practices to which it was meant to apply together into a coherent possibility-for-being that then became more publicly and widely available. By doing so, being “laid back” also became a possibility and style that was up for contestation, elaboration, and modification through new practices, what people said about them, whether the term could be used for certain actions, certain people, etc. In fact, previously established possibilities began to show up for people in the light of being “laid back.” That is, many people began to evaluate jobs, decisions, places of living, etc., based on whether they would afford a certain degree of being “laid back.” The phrase was not arbitrary because it creatively and skillfully used words that drew upon other meanings and connotations already available in our culture and successfully applied them to a new realm. It “stuck” because it aptly brought into focus the meaning of the style latent in the practices to which it applied. It focused a style to which more and more people could meaningfully relate.

Though this example was not one of theoretical or governmental classification, it can easily serve as a model for them. It shows how practices, terms, identities, kinds, and descriptions can all interact and be responsive to each other. On the other hand, it can also point toward what can go wrong with government and theoretical classification. As Hacking sometimes alludes in his writings on human kinds, government and theoretical classification can often serve as mechanisms for control and oppression of marginalized kinds in human culture and society. In contrast to invention of the term “laid back,” which may be said to be the result of being in tune

18 Dreyfus’s remarks are rather brief. The rest of the discussion is my elaboration on his brief remarks. However, they are influenced by other remarks he has made on similar issues.
with and receptive to the style and meaning of the practices involved, government and theoretical classifications often serve politically or ideologically driven motives. Not only, then, might we have reason to doubt the accuracy of governmental and theoretical classifications, there may be ethical reasons to be concerned with their classificatory practices (and the kinds they help create through subsequent looping effects). That is, governmental and classificatory classifications may be “sticky” only in so far as they are coercively so. This leads me to one last point of Hacking’s and an example of my own. They should serve to demonstrate the point about ethics being involved in the feedback loops as well as continue to illustrate the ways in which terms, descriptions, and their dissemination in society influence human kinds.

The Ethics of Human Kinds

In some of his writings in the study of making up people, Hacking has suggested that, for human kinds, there may be a certain “indeterminacy of the past” (e.g., *Rewriting the Soul* and “Kind-making: The Case of Child Abuse”). The idea is that new kinds and human actions that were not available in the past may be used to reinterpret the past. The question is whether such past actions were really of the new kind or not. Hacking asks, for example, whether certain cases from history that would probably be considered child abuse today were really so when they occurred. He also asks whether it is appropriate to apply certain psychological disorders like Post Traumatic Stress Disorder (PTSD) to historical figures, in some cases excusing them of wrongdoing (see “Indeterminacy in the Past: On the Recent Discussion of Chapter 17 of *Rewriting the Soul*”). He has suggested that there is no truth to the matter, that some things in the past are ambiguous and indeterminate due to future developments. This claim has engendered quite a debate recently in the journal *History of the Human Sciences* (see “Indeterminacy in the Past”; “Indeterminacy in the Past: On the Recent Discussion of Chapter 17 of *Rewriting the Soul*”; McMillan’s “Under a Redescription”; and Leudar and Sharrock’s “Essay Review: Multiplying the Multiplicity: Are Dissociative Identity Disorders ‘Real?’”). My aim is not to try to sort out that whole issue here. What I
think Hacking does show, though, is that another unique attribute of humans and their kinds is that their pasts are something toward which they can meaningfully comport—those pasts can serve as motives for the sake of existentially meaningful action. This seems to be a significant property of human beings. Only they can reinterpret their past and have it be a motive for action and creating new possible kinds. People’s pasts are a significant source of existential meaning for them. In reflecting on this idea, I remembered the following incident.

It is a well-known principle of introductory economics that “sunk costs” are irrelevant for future economic choices. That is, people and businesses should only base their economic decisions on the most profitable alternative and not on past expenses. Having “sunk” a lot of money into an unprofitable venture should not keep someone from cutting their losses and taking up a better option. I was once in an economics course where the professor tried to teach (not just illustrate) the students to use this principle in their personal relationships. History should not matter, he said, to whether you continue a relationship with someone. The only thing that matters is which currently possible relationships are likely to bring the most satisfaction in the future. Whether or not this particular piece of advice is good, it strikes me that this is an example of a theoretical social science extending its descriptions for some kinds of economic actions into further realms of human action and identity. Currently, economics and its ideology seem to enjoy particular sway in our society. I wonder, for example, if the influence of the economic picture of humans could be such that it could seriously undermine unique properties of human kinds. Whatever its scope, surely there is something worthwhile about the unique ability of humans to draw upon the meanings in their past as motivations to act on for the future. Could it be possible that intrusion of economic principles onto the way humans define themselves could seriously alter or even effectively disable some of their unique properties that have here been uncovered using Hacking’s approach to their study? It may not seem likely, but here is where the importance of the unpredictability of human kinds comes in. Throughout his writing and research, Hacking has shown how quickly and unpredictably human kinds have been and can be “made up.” He shows how certain invented statistical categories shoot up from near
zero to hundreds of thousands in the matter of a few years. I think it is a legitimate and serious question whether the social sciences can significantly affect some of the most important aspects of what it means to be human. Thus, to the extent that human sciences have the power to alter some of our unique abilities, it becomes partly an ethical question what kinds of humans we want to be and whether and to what extent the human sciences should interfere with the development of whatever kinds we may decide are most worth pursuing. It may just be partly an ethical question whether certain human sciences should be restricted from doing some of their work or extending their theorizing in certain ways. This is a dark and foreboding thought. Many historical attempts at the supposed ethical control of science have resulted in failure or even worse kinds of oppression and abuse. It is not within the scope of this paper to fully address this issue. I have merely raised it to illustrate the role that ethics and politics may have in the making up of human kinds.

Conclusion

In this paper I have defended Hacking’s work on human kinds by clarifying and elaborating on some of its key elements. Specifically, I have argued that the role of existential meaning in mediating the feedback between cultural practices and classificatory practices is part of what makes human kinds different from natural kinds. Human kinds differ from natural kinds in that their epistemology, semantics, and ethics feed back onto their own ontology in ways different from natural kinds. I have illustrated ways in which this is done and have begun to sketch out a more detailed theory of how terms, descriptions, actions, and identities all interact in the formation of feedback loops. To do this I have argued that there are ongoing “micro-level” feedback loops that are crucial in the creation of kinds that are identified at the larger, theoretical level. I have only given a rough sketch of what these “micro-loops” might look like and how they might operate. I think that further research on them is warranted. Furthermore, I have argued that, while my conclusions do not entail skepticism about general social knowledge, they do imply certain limits to theoretical approaches to human kinds that are modeled on the methods of the
natural sciences. There is a possibility that pursuing certain lines of research into the study of human kinds would violate the accommodation thesis of Boyd, shifting important causal structures of human kinds sufficiently enough to pass off at least partially creative results of social sciences as pure discovery. This possibility further entails the need for ethics in the study of human beings. Finally, it may be responded to my arguments that for all this, it does not entail that human beings are not natural. That may certainly be the case. The purpose of this paper has not been to argue over how to define or apply the term “natural.” Rather, it has been to show important differences between human and other kinds—differences that matter in philosophical discussion about essence and ontology in relation to these kinds.  

19 I would like to thank the editors of Aporia for useful feedback that improved the quality of the essay.
Works Cited


Alvin Plantinga addresses the classic ontological argument in two books published in 1974: *The Nature of Necessity* and *God, Freedom, and Evil*. In each of these books, he analyzes the classical ontological argument and eventually formulates his own contemporary modal version. In *The Nature of Necessity*, the argument is presented in a technical form, using many of the concepts Plantinga developed throughout the book. *God, Freedom, and Evil* contains a streamlined, more easily read version of the argument, but it omits some of the underlying subtle technicalities that explain Plantinga’s formulation. However, *God, Freedom, and Evil* contains a more thorough analysis of objections to the ontological argument, namely those of Gaunilo and Kant. In this paper, I will discuss Plantinga’s contemporary modal version of the ontological argument and will draw from the sources discussed above. I will then argue that Plantinga must simultaneously avoid two conflicting problems in order to formulate a successful argument. Specifically, Plantinga must formulate his argument to avoid both his own criticism of St. Anselm’s original argument and his own criticism of Gaunilo’s objection from “On Behalf of the Fool.” However, since Plantinga cannot simultaneously avoid both these problems, his contemporary modal version of the ontological argument is unsuccessful.

*Tyrel Mears is a senior majoring in philosophy at Lewis and Clark College in Portland, Oregon. After graduating he plans on living in Portland and applying to graduate programs in philosophy.*
I. Background

1.1 St. Anselm

Let us begin with Plantinga’s take on St. Anselm’s classic argument:

(1) God exists in the understanding but not in reality.

(2) Existence in reality is greater than existence in the understanding alone.

(3) God’s existence in reality is conceivable.

(4) If God did exist in reality, then he would be greater than he is [from (1) and (2)].

(5) It is conceivable that there be a being greater than God [(3) and (4)].

(6) It is conceivable that there be a being greater than the being than which nothing greater can be conceived [(5) by the definition of “God”] (Plantinga, The Nature of Necessity 198).1

But surely

(7) It is false that it is conceivable that there be a being greater than the being than which none greater can be conceived.

Because statements (6) and (7) contradict each other, we may conclude that

(8) It is false that God exists in the understanding but not in reality.

It seems that we can conclude that since God clearly exists in the understanding, then he must also exist in reality. There are at least two objections to this argument, namely those of Gaunilo and Kant.

1 All future references to works by Plantinga will use the name of the work followed by the page number.
1.2 GAUNILO’S OBJECTION

On Plantinga’s reading, Gaunilo held that absurd objects can be formulated as objects “than which none greater can be conceived”; therefore, if existence is a great-making property, this type of argument proves that these absurd objects must exist (God, Freedom, and Evil 89–90). Specifically, Gaunilo claimed to be able to prove the existence of an island than which none greater can be conceived (God, Freedom, and Evil 89–90). He could do this simply by replacing “God” with “the greatest island” and “being greater than which none greater can be conceived” with “island greater than which none greater can be conceived.” Then, if existence is a great-making property for islands, there must exist an island than which none greater can be conceived. But the existence of such an island is absurd. Gaunilo’s point is that this method of argument is not valid since it clearly produces false conclusions from premises Anselm should accept.

Plantinga discredits this objection by noting that islands (and things of this nature) are qualified as great according to properties that have no intrinsic maxima (God, Freedom, and Evil 91). Properties that do not exhibit intrinsic maxima cannot ever be had to a maximum degree; so, things that have great-making properties that do not exhibit intrinsic maxima cannot ever be maximally great. Therefore, “the idea of a greatest possible island is an inconsistent or incoherent idea; it’s not possible that there be such a thing” (God, Freedom, and Evil 91). Thus, if an object \( x \) has a great-making property \( P \) that does not have an intrinsic maximum, we cannot coherently conceive of the greatest possible \( x \) since there is no maximum of \( P \) that \( x \) may have. Plantinga then concludes that premise (3) cannot be formulated in terms of Gaunilo’s island or anything that has a great-making property without an intrinsic maximum. However, he does not think that this criticism applies to Anselm’s argument (God, Freedom, and Evil 90–91).

---

2 Plantinga notes that Gaunilo’s original objection does not follow St. Anselm’s argument exactly. Instead, he speaks of an island that is in fact greater than any other; however, to be parallel Gaunilo should have spoken of an island than which none greater can be conceived. I have transformed his following objection accordingly.

3 Plantinga does not give a formal definition for the term “intrinsic maximum” when he introduces it here, even though it is an essential concept in his eventual argument. For the time being, it will suffice to know that Plantinga is using “intrinsic maximum” in the following way. A property \( P \) has an intrinsic maximum if and only if there is a degree of \( P \) such that it is impossible to have more of \( P \). I give a formal definition section 3.2 of this paper.
Plantinga reasons that great-making properties of the greatest possible being do have intrinsic maxima. Specifically, Plantinga considers knowledge, power, and moral excellence as great-making properties. He thinks that all of these properties have intrinsic maxima, but Plantinga also believes that there may be other great-making properties (for example, love) where it is unclear whether or not they have intrinsic maxima (God, Freedom, and Evil 91). So, Plantinga concedes that the argument may have a weakness insofar as the greatest possible being might have great-making properties that do not have intrinsic maxima (God, Freedom, and Evil 91). If this was true, then the ontological argument would be subject to Plantinga’s criticism of Gaunilo’s objection. That is, if the greatest possible being had great-making properties without intrinsic maxima, then the greatest possible being would be inconsistent or incoherent, just like Gaunilo’s island. So, Plantinga concludes that the argument needs further attention (God, Freedom, and Evil 91).

Plantinga leaves us with the impression, though, that the ontological argument can escape Gaunilo’s objection.

1.3 Kant’s Objection

Kant gave two famous objections to the ontological argument in his Critique of Pure Reason. Plantinga thinks that Kant’s first objection is that “no existential proposition—one that asserts the existence of something or other—is necessarily true; the reason, he says, is that no contra-existential (the denial of an existential) is contradictory or inconsistent” (God, Freedom, and Evil 92–93). Plantinga dismisses this objection since, in his opinion, Kant gives no argument to support it and since Plantinga cannot think of a reason to accept it (God, Freedom, and Evil 93–94). In fact, Plantinga argues that denials of existentials (singular negative existentials) are sometimes true (The Nature of Necessity 151). For example, note that (1) may be

---

4 Plantinga says we should note this weak point but does not want to discuss it and suggests we move on. This point deserves more attention. For example, Patrick Grim has argued that knowledge has no intrinsic maximum since there is no set of all truths. This claim is based upon the idea that if there was a set of all truths then the power set of all truths will have as its members subsets of the set of all truths, which will have corresponding truths. But by Cantor’s power set theorem, the power set of any set is larger than the original set. Thus, there is no set of all truths. Therefore, there cannot be an omniscient being. This topic may have some interesting connections to what I bring up in this paper. For more information, please see Grim’s works “The Being that Knew Too Much” and “Truth, Omniscience, and Cantorian Arguments: An Exchange” (hereafter cited by title).
construed as: “It is possible that −(God does exist).” That is, (1) is asserting that that a singular negative existential proposition “−(God does exist)” is possibly true. However, this negation may be interpreted in two different ways. Consider the predicative proposition:

(a) God has the property of nonexistence

versus the impredicative proposition

(b) God does not have the property of existence.

There is a clear difference here (The Nature of Necessity 149–152).⁵ According to Plantinga, we may sensibly assert impredicative singular negative existential propositions but not predicative singular negative existential propositions (The Nature of Necessity 151). This is because if we take (1) as (a), then we are predicating a property (nonexistence) of something that does not exist. But it is absurd to assign properties to objects that do not exist (The Nature of Necessity 152). On the other hand, (b) is just a variant of “It is false that God has the property of existence” (The Nature of Necessity 151). Thus, according to Plantinga, the correct (and only sensible) way of taking (1), which is the negation of a singular existential proposition, is the impredicative (b). This means that existence is an essential property of any object of which we may predicate properties (The Nature of Necessity 152).⁶ So, it is possible to construe (1) as (b) and not predicate any properties.

Kant’s second objection is that existence is not a real predicate; that is, one cannot predicate the existence (or nonexistence) of any object in any sensible way (Kant, Critique of Pure Reason, 567).⁷ Kant thought that the predicate “exists” is already contained within the subject of a proposition. If this is true, then it doesn’t make sense to claim that “existence is a great-making property” (Kant 567–68). Plantinga interprets this objection: “Perhaps Kant means to make a point that we could put saying that it’s not possible to define things into existence” (God, Freedom, and Evil 95). This seems

---

⁵ This is the section where Plantinga introduces and discusses the difference between predicative and impredicative statements and how this distinction applies to singular negative existential propositions.

⁶ This is not to say that any object of which we may predicate properties necessarily exists. This is explained in the following paragraph.

⁷ This work will be cited by author’s last name only. For discussion of the ontological argument, see Kant 563–69. For discussion of conceptual existence in general, see Kant 325–26. See also God, Freedom, and Evil 93–97 for discussion of specific passages from Critique of Pure Reason.
like a reasonable claim; any procedure that "defines things into existence" must be flawed. Plantinga gives an example of the kind of argument he takes Kant to be criticizing: suppose we define two terms, "bachelors" and "superbachelors":

\[ x \text{ is a bachelor if and only if } x \text{ has } P_1, P_2, P_3, \ldots, P_n \]
\[ x \text{ is a superbachelor if and only if } x \text{ has } P_1, P_2, P_3, \ldots, P_n, \text{ and } x \text{ exists (God, Freedom, and Evil 96).} \]

In these definitions, \( P_1, P_2, P_3, \ldots, P_n \) are essential properties of a bachelor such as "being unmarried, being male, being over the age of twenty-five, and the like" (God, Freedom, and Evil 95). Note that superbachelors have all the essential properties of the bachelors plus the essential property of existence (God, Freedom, and Evil 95). It is tempting to say "Necessarily superbachelors exist" (God, Freedom, and Evil 95). But this is incorrect. Instead, what we should say is "Superbachelors exist necessarily or essentially." From this it does not follow that there must be superbachelors, but that if there are superbachelors, then they exist. However, this latter conclusion yields little cognitive significance. In fact, adding the predicate of existence does not affect our definition of bachelors since all bachelors are necessarily superbachelors and vice versa (God, Freedom, and Evil 97). Therefore, Plantinga accepts that one cannot define things into existence. He does not, however, think this claim has any affect on the ontological argument (God, Freedom, and Evil 97–98).

1.4 Plantinga’s Objection

Plantinga himself objects to St. Anselm’s version of the argument. Specifically, he sees a problem with (6). He thinks that it is not necessarily false, as (7) asserts (The Nature of Necessity 202–213). The argument is stated in terms of possible worlds in the following (flawed) way:

(9) God does not exist in the actual world.

(10) For any worlds \( W \) and \( W^* \) and object \( x \), if \( x \) exists in \( W \) and \( x \) does not exist in \( W^* \), then the greatness of \( x \) in \( W \) exceeds the greatness of \( x \) in \( W^* \).
(11) It is possible that God exists.

(12) So there is a possible world $W$ such that God exists in $W$ [from (11)].

(13) God exists in $W$ and God does not exist in the actual world [from (9) and (12)].

(14) If God exists in $W$ and God does not exist in the actual world, then the greatness of God in $W$ exceeds the greatness of God in the actual world [from (10)].

(15) So the greatness of God in $W$ exceeds the greatness of God in the actual world [(13) and (14)].

(16) So there is a possible being $x$ and a world $W$ such that the greatness of $x$ in $W$ exceeds the greatness of God in actuality [(15)].

(17) So it is possible that there be a being greater than God is [(16)].

(18) Hence it is possible that there be a being greater than the being than which it is not possible that there be a greater [from (17) by definition of “God”].

(19) It is not possible that there be a being greater than which it is not possible that there be a greater (The Nature of Necessity 202).

Thus, (9) is reduced to a contradiction. However, let us look more closely at the supposed contradiction. Plantinga claims that we should take (12), (18), and (19) differently since it is unclear which greatness (relative to a world) that we are referring to in these propositions. Thus, he claims the correct way is as follows:
There is a possible world \( W \) such that the being whose greatness in some world is nowhere exceeded, exists in \( W \).

There is a possible world \( W \) and a possible being \( x \) such that the greatness of \( x \) in \( W \) exceeds the greatness of \( x \) in the actual world.

There is no possible world \( W^* \) and being \( x \) such that the greatness of \( x \) in \( W^* \) exceeds the greatness of God in \( W \) \((The\ Nature\ of\ Necessity\ 203-5)\).

Note that (19*) does not contradict (18*) the way that (19) contradicts (18). What this amounts to is that (18) should correctly claim that there exists a being that is greater than God in the actual world. And (19) should correctly claim that "It’s not possible that there be a being whose greatness surpasses that enjoyed by the unsurpassably great being in the worlds where its greatness is at a maximum" \((God,\ Freedom,\ and\ Evil\ 103)\). Since the greatness of this being varies from world to world, the apparent contradiction of "there exists a being greater than the being than which it is not possible that there be a greater" is only a surface contradiction. Thus, this reductio version of the argument fails since it does not produce a formal contradiction. But all is not lost. The argument does allow us to conclude that this being exists in at least one possible world; however, we have no reason to think this world is the actual world \((God,\ Freedom,\ and\ Evil\ 103-4)\).

II. Plantinga’s Modern Modal Ontological Argument

Plantinga has now dealt with Gaunilo’s objection and tentatively with Kant’s. However, he needs to formulate his argument so that it is not subject to his criticism of St. Anselm’s argument. Plantinga’s first step is to draw a distinction between greatness that is world-dependent and greatness that is universal. These are defined as “excellence” and “greatness,” respectively \((The\ Nature\ of\ Necessity\ 214)\). In this manner, excellence is calculated by the great-making properties a being has in a particular world, while greatness is calculated from the world-dependent property of excellence. Plantinga defines and uses these terms as follows:
Those who are fond of the calculus might put it by saying that there is a function assigning to each being in each world a degree of excellence; and a being’s greatness is to be computed (by someone unusually well informed) by integrating its excellence over all possible worlds. Then it is plausible to suppose that the maximal degree of greatness entails maximal excellence in every world. A being, then, has the maximal degree of greatness in a given world $W$ only if it has maximal excellence in every possible world. But maximal excellence entails omniscience, omnipotence, and moral perfection (God, Freedom, and Evil 107–8).

All this greatly affects the definition of God as the greatest possible being. From the above passage it is clear that for God to be maximally great he must have maximal excellence in every possible world, which entails that he exists in every possible world. So, according to this definition, necessary existence is an essential property of God. That is, if God exists then he necessarily exists, necessarily or essentially. Plantinga can now reformulate the argument:

(25) It is possible that there be a being that has maximal greatness.

(26) So, there is a possible being that in some world $W$ has maximal greatness.

(27) A being has maximal greatness in a given world only if it has maximal excellence in every world.

(28) A being has maximal excellence in a given world only if it has omniscience, omnipotence, and moral perfection in that world (God, Freedom, and Evil 108).

From (26)–(28), we can conclude that there is a being that enjoys maximal excellence (i.e., has omniscience, omnipotence, and moral perfection) in every possible world, including the actual world. But a being cannot enjoy maximal excellence in a world if it does not exist in that world (God, Freedom, and Evil 108). Therefore, a being that is maximally excellent in every possible world, i.e., the being that is maximally great, exists in every
possible world, including the actual world. This argument essentially differs from previous versions in that it is not a reductio. It may be construed as one if we add the reductio hypothesis that God does not exist in the actual world, but this extra step is unnecessary. All that is required here is the first premise (25) and the definition of God as the greatest possible being (God, Freedom, and Evil 109).

Plantinga’s argument is almost complete at this point. The last issue is that of possible but unactual objects, such as the object in (25). Plantinga writes that he is “inclined to think the supposition that there are such things—things that are possible but don’t in fact exist—is either unintelligible or necessarily false” (God, Freedom, and Evil 110). That said, he reformulates the argument in terms of properties being “exemplified” or “instantiated” in a world. Consider:

(29) There is a possible world in which maximal greatness is instantiated.

(30) Necessarily, a being is maximally great only if it has maximal excellence in every world.

(31) Necessarily, a being has maximal excellence in every world only if it has omniscience, omnipotence, and moral perfection in every world (God, Freedom, and Evil 111).

So, if the property of maximal greatness is possible, then it is impossible that a being that exemplifies maximal greatness (which entails maximal excellence in every possible world) fails to exist necessarily.

8 See also Chapter VIII of The Nature of Necessity for arguments defending this position.
9 Plantinga seems to think that this makes Kant’s objection completely irrelevant. However, (30) may still raise some suspicion. He supports himself by saying that we should not be any more concerned about (30) than we are about “Necessarily, a thing is a unicorn only if it has one horn.” See God, Freedom, and Evil 111.
III. Preliminary Concerns

3.1 Existence and greatness

This argument essentially contains three pieces:

(a) The definition of the property “maximal greatness”

(b) The possibility that maximal greatness is instantiated

(c) “Possibly (necessarily p)” is equivalent to “necessarily p.”

It is important to note that most of the machinery of this argument is contained within (a), Plantinga’s definition of the property of maximal greatness. Appropriately, most of the reasoning behind the argument was about (a). Observe that (c) is simply a rule of modal logic that allows an inference from (a) and (b); however, arguments for or against (c) should not be dependent on our current debate. At this point, I am willing to accept (c) and think that Plantinga is correct in his use of it. Unlike (c) however, whether or not I am willing to accept (b) is entirely dependent on what I think about (a).

What exactly does the concept of maximal greatness entail? What makes something great? Recall that at one point Plantinga claimed existence was a great-making property (God, Freedom, and Evil 98–100). Suppose that \(x\) exists in \(W\) and \(x\) does not exist in \(W^*\). Then, Plantinga claims that the greatness of \(x\) in \(W\) exceeds the greatness of \(x\) in \(W^*\). Greatness is measured in degrees and so is expressed in values of degree. Therefore, a comparison of two beings’ greatesses is a comparison of two values. Now, I agree with Plantinga that an object \(x\) does not have any properties in worlds in which it does not exist; so that if \(x\) does not exist in \(W^*\) then \(x\) does not have a property of greatness (of any value) in \(W^*\) (The Nature of Necessity 152). Suppose there is a function that computes the value of an object’s greatness in a world; let \(f\) be such a function. If \(f\) is a

10 This is an axiom in the S5 system of quantified modal logic that is often a topic of debate.
11 During the following, greatness is used as it was before the distinction between “excellence” and “greatness” was made.
12 This is also mentioned in (10) above.
function, then \( f(x \in W) = \) some value. However, \( f(x \in W^*) \) doesn’t have any value, even a value of zero. The \( x \in W^* \) that we are trying to use in function \( f \) is non-referring. If we were to assign any value to \( f(x \in W^*) \), then we would be assigning a property to \( x \in W^* \), namely the property of having a certain value of greatness. But \( x \) does not exist in \( W^* \). Now, what Plantinga wants to be able to say is that \([f(x \in W)] > [f(x \in W^*)]\). But this requires a comparison. Attempting to do this comparison is akin to giving a person two pieces of paper, one with a number on it and one without a number on it, and asking “Which piece of paper has a greater number on it?” This is not a legitimate question to ask since it presupposes that there are two numbers to compare. The proposition “The greatness of \( x \) in \( W^* \) is 0” is false in the same way that “The present king of France is bald” is false (Russell 167–180). However, what we are dealing with here is not the ability to evaluate the truth value of some proposition but rather the ability to make a sensible comparison between two (numerical, or at least orderable) values. Note that \([f(x \in W)] > [f(x \in W^*)]\) entails that \( f(x \in W^*) = n \) is true for some \( n \); but for any value of \( n \), “\( f(x \in W^*) = n \)” is false. So the claim made by the consequent of (10) rests on a comparison that is impossible to make if the antecedent conditions of (10) are met.

Soon after introducing (10) above in *The Nature of Necessity*, Plantinga amends it to:

\[
(10^*) \text{ For any world } W \text{ and object } x, \text{ if } x \text{ does not exist in } W, \text{ then there is a world } W^* \text{ such that the greatness of } x \text{ in } W^* \text{ exceeds the greatness of } x \text{ in } W \text{ (*The Nature of Necessity 203).}^{13}
\]

Plantinga claims that this weaker premise will serve the same function in his argument. This is true. Nonetheless, this weaker premise is also flawed. Suppose that \( x \) does not exist in any world. Then \( x \) would not exist in \( W \). It does not follow that there is a world \( W^* \) in which the greatness of \( x \) exceeds the greatness of \( x \) in \( W \), since \( x \) does not exist in any world.\(^{14}\) The phrase

\(^{13}\) Note that there is a switch of convention here; now \( x \) does not exist in \( W \) but \( x \) does exist in \( W^* \). I do this to stay true to the convention Plantinga uses in the section that I am discussing so that any reference to the section will reflect the convention I use while discussing it.

\(^{14}\) This is an example where it seems that the antecedent conditions of (10*) are met, but the consequent does not follow. But, there may be a subtlety in the antecedent such that to call \( x \) an “object” it must be the case that \( x \) exists in at least one world, namely \( W^* \).
“the greatness of $x$ in $W$” is non-referring as it is used in (10*). Also, it is not clear whether “the greatness of $x$ in $W^*$” is referring or not. However, the question arises as to whether Plantinga was implying that $x$ must exist in at least one world ($W^*$) by naming $x$ as an object. If so, then $x$ does exist in $W^*$, but this turns (10*) into (10). Now, having criticized these claims, we should observe that Plantinga does not end up using them explicitly in his final argument. But is one of them wrapped up in the definition of “maximal greatness”? No, at least not in his eventual definition that is used. Maximal greatness is maximal excellence in every world. But maximal excellence in any world requires (entails) existence in that world. It is not the case that existence is a great-making property but rather is a necessary condition of maximal greatness (The Nature of Necessity 214). So, neither (10) nor (10*) are essential to the final argument.

3.2 Concept of Intrinsic Maximum

Plantinga introduces the concept of an “intrinsic maximum” in order to refute Gaunilo’s original objection. He never offered a formal definition; however, he did write:

The qualities that make for greatness in islands—number of palm trees, amount and quality of coconuts, for example—most of these qualities have no intrinsic maximum. That is, there is no degree of productivity or number of palm trees (or of dancing girls) such that it is impossible that an island display more of that quality (God, Freedom, and Evil 91).

So, intrinsic maxima are properties of qualities (or properties) of an object. What Plantinga seems to be suggesting is a definition of “intrinsic maximum” as follows:

A property $P$ has an intrinsic maximum if and only if there exists a degree (or number) $m$ of $P$ such that for any degree (or number) $n$ of $P$, $n < m$ or $n = m$.

Considering how central this concept is to Plantinga’s argument—he uses it to both reject Gaunilo’s objection and validate (b)—a formal definition is

15 The qualities that are of specific interest to us are great-making qualities.
long overdue. An important distinction needs to be made here between a maximum and a limit. Consider \([0, 10)\) and \([0, 10]\), sets of real numbers.\(^{16}\) Note that while \([0,10)\) has an upper limit of 10, it does not exhibit a maximum since for any number \(n\) on the interval there exists a number \(y\) on the interval such that \(y > n\). On the other hand, \([0, 10]\) does exhibit a maximum: let \(m = 10\); note that for any \(n\) on \([0, 10]\), \(n < m\) or \(n = m\).

How does this apply to properties and intrinsic maxima? Keep in mind that not all properties, such as the property of being taller than six feet, admit of degrees; however, of those that do there are two kinds: those with intrinsic maxima and those without intrinsic maxima. Suppose property \(P\) is a property that does admit a degree (i.e., property \(P\) can be had by an object in quantifiable degrees). Now for an object \(x\), if \(x\) has \(P\), then \(x\) has a certain value of \(P\). That is to say, if an object \(x\) has property \(P\), then \(x\) has some value of \(P\) that is a member of some certain interval of possible values of \(P\). The nature of this interval determines whether or not \(P\) has an intrinsic maximum. Let’s take the property of mass as an example. Suppose an object \(x\) has the property of mass. If so, then \(x\) is a physical object. This implies that \(x\) has a positive value of mass. So, the value of mass that \(x\) has is a member of a set that has a lower limit of zero and does not have zero as a member. Furthermore, for the sake of the example let’s assume that there is no limit to the amount of mass an object may have. Then the value of mass \(x\) has falls somewhere between 0 and infinity. For this example, the set of possible values of mass does not have an upper bound, so we may infer that mass does not have an intrinsic maximum.\(^{17}\) Note that if a property has an intrinsic maximum, then it has an upper limit. However, the converse of this is not true. Furthermore, note that if a property \(P\) has an intrinsic maximum \(m\), then it is impossible for any object \(x\) to have more than \(m\) of \(P\). So, if a property has an intrinsic maximum, then it has that intrinsic maximum necessarily or essentially. This means that the property of having an intrinsic maximum is not world-dependent property.

\(^{16}\) The difference between \([0,10)\) and \([0,10]\) is that 10 is not a member of the former, but 10 is a member of the latter set. Note that 0 is a member of both.

\(^{17}\) As used here, upper bound is just another way of saying upper limit.
Greatness is “to be computed (by someone unusually well informed) by integrating [a being’s] excellence over all possible worlds” (God, Freedom, and Evil 107–8). Let’s not push Plantinga’s calculus analogy; indeed, I don’t think it can be pushed at all without falling apart (Fitzpatrick, Advanced Calculus 118). Rather, what I think we should concern ourselves with is what Plantinga intended to suggest by this statement. If I understand him correctly, we could calculate the greatness of a being by a simple addition of its excellence in each world. Thus, let $h$ be the excellence function such that

$$h(B_i) = \text{excellence of being } B \text{ in world } W_i.$$ 

Let $g$ be the greatness function such that

$$g(B) = h(B_1) + h(B_2) + h(B_3) + \ldots + h(B_n) \text{ as } n \text{ approaches infinity.}$$

Greatness has now been properly defined as a calculation (by addition) of a being’s excellence in all worlds in which that being exists.

IV. Criticism of Plantinga’s Argument

4.1 Two Conflicting Requirements for a Successful Ontological Argument

Recall that Plantinga remedies his own objection (section 1.4) by building necessary existence into the definition of maximal greatness. That is to say, he defined the property of maximal greatness such that if maximal greatness is possibly instantiated, then it is necessarily instantiated.

---

18 Hereafter simply “Fitzpatrick” followed by the page number. Whether or not it is possible to integrate this function depends how we would technically formulate the function of excellence. Any attempt at this procedure would be troublesome to say the least. For example, suppose that a being $B$ exists in only one world. It seems, then, that there would only be one point representing the being’s excellence on the graph of this being’s greatness. But if this is so, then the being in question would not have any greatness according to the integration analogy. For a criterion of integrability, see Fitzpatrick.

19 Note that this procedure does not run into any problems in worlds in which the being does not exist. In such worlds, the being does not have excellence, thus not adding to the being’s greatness.

20 Here and throughout the rest of my paper I am using the word “function” non-technically.
Plantinga was forced to do this to avoid his own criticism of St. Anselm’s argument. However, before he offered his criticism he considered Gaunilo’s objection, “On Behalf of the Fool.” In response, Plantinga claimed that objects that have as their great-making properties those which do not have intrinsic maxima cannot ever attain a maximal degree of greatness. Thus, the idea of a greatest possible island or, more correctly, the property of a maximally great island, is inconsistent or incoherent. Since these properties are inconsistent or incoherent, they are not possible and thus not possibly instantiated. This conclusion allows Plantinga to avoid accepting the existence of absurd objects like Gaunilo’s island. The reason is that (b) above cannot be formulated in terms of the property of the greatest island or in terms of any property that does not have an intrinsic maximum. The property of maximal greatness (as seen in (b)) is safe from Plantinga’s reply to Gaunilo since maximal greatness has an intrinsic maximum. Is this true? Does maximal greatness have an intrinsic maximum? Plantinga wrote:

Anselm clearly has in mind such properties as wisdom, knowledge, power, and moral excellence or moral perfection. And certainly knowledge, for example, does have an intrinsic maximum: if for every proposition $p$, a being $B$ knows whether or not $p$ is true, then $B$ has a degree of knowledge that is utterly unsurpassable. So a greatest possible being would have to have this level of knowledge: it would have to be omniscient. Similarly for power; omnipotence is a degree of power that can’t possibly be excelled. Moral perfection or moral excellence is perhaps not quite so clear; still a being could perhaps always do what is morally right, so that it would not be possible for it to be exceeded along those lines. . . . And what about the relevant qualities here—love, or acting out of love: do they have intrinsic maxima? The answer isn’t very clear either way. Rather than pause to discuss this question, let’s note simply that there may be a weak point here in Anselm’s argument and move on (God, Freedom, and Evil 91).21

Plantinga is claiming that his argument is not subject to his own criticism of Gaunilo’s objection since knowledge, power, moral perfection, etc., have

---

21 I don’t think he ever returns to this point, and he never gives justification for moving on. So this is a weak point in his final argument as well. See footnote 12 for why this point may also be reason to reject Plantinga’s argument.
intrinsic maxima. That is to say, greatness has an intrinsic maximum so that it is consistent or coherent to assert (b), which is about maximal greatness. But these virtues (knowledge, power, moral perfection, etc.) are qualities of excellence. So, what we can infer from the above argument (if it is indeed correct) is that excellence has an intrinsic maximum. What this means, according to our definition of "intrinsic maximum," is that the set of possible values of excellence has a closed upper bound—that is, the upper limit is a member of the set. This implies that there is a possible value of excellence e such that for any value of excellence n, n < e or n = e. Now, I think it is safe to assume that Plantinga would agree e is positive. What we cannot infer, at least directly, is that greatness has an intrinsic maximum. The being that exemplifies maximal greatness will have maximal excellence in all worlds. This excellence, although undoubtedly very great, has a value of e that is a finite positive value. However, when we add this constant positive value from all possible worlds (an infinite amount), we cannot possibly be left with a finite value, thus implying that there is no intrinsic maximum of greatness. What Plantinga did establish is that excellence has an intrinsic maximum, but he never established that greatness has an intrinsic maximum after he defined "greatness" as an addition (or less accurately, as an integration) of excellence in all possible worlds.

Here Plantinga may very well say that greatness does have an intrinsic maximum since "The limiting degree of greatness, therefore, would be enjoyed in a given world W only by a being who had maximal excellence in W and in every other possible world as well" (The Nature of Necessity 214, my emphasis). Admittedly, this seems reasonable. However, it appeals to confusion between what constitutes a limit and what constitutes a maximum. Let's grant Plantinga that the "limiting degree of greatness" entails maximal excellence in every world. Does there exist some value m of greatness such that for any value n of greatness, n < m or n = m? The correct response is no. An answer of yes would be equivalent to making the mistake of saying that since [0,10) has a limit of 10, there exists a number of [0,10) that is

---

22 See footnote 12 for Patrick Grim's related argument.
23 Whether or not excellence has a lower bound and, if it does, whether or not it is zero or a negative value may be debated. But we need not enter this debate here since the being we are interested in (the greatest possible being) has the maximal level of excellence, e, in every possible world.
24 I am not ready to accept that there is even a limiting degree of greatness; however, failing to exhibit an intrinsic maximum is sufficient to reject the argument.
the greatest. However, I claim that greatness does not even have an upper bound (or a limiting degree). Consider the greatness function of the greatest possible being $B$:

$$g(B) = h(B_1) + h(B_2) + h(B_3) + \ldots + h(B_n)$$

as $n$ approaches infinity, where $h(B_i)$ is the excellence of being $B$ in world $W_i$.

Note that $h(B_i) = e$, where $e$ is the intrinsic maximum value of excellence and $i$ is any natural number. The greatness function of the greatest possible being, $g(B)$, is an infinite series starting at one and summing the value of excellence in every possible world. Since $h(B_i) = e$, for all $i$ that are natural numbers, $g(B)$ is an infinite summation of the constant positive value $e$. Note that $g(B)$ is not bounded since there is no number $m$ such that $g(B) < m$ or $g(B) = m$ for every natural number. Thus, the infinite series $g(B)$ diverges (Fitzpatrick 27). This implies that there is no maximum value of greatness.

So asserting (b) above is akin to asserting “It is possible that there is a natural number greater than all other natural numbers.”

There is something curious about the greatest possible being having the maximal degree of excellence, $e$, in every possible world. Excellence is a world-dependent property. However, the maximal degree of excellence is not world-dependent. But it seems that there might be a limit to how excellent a being could be in some world that is less that the maximal degree of excellence $e$. Suppose that we take Plantinga’s example of moral excellence as an excellence-making property. It seems that in a world in which no other beings exist, the greatest possible being could not have much in terms of moral excellence. Could it be that it is impossible for any (even the greatest possible) being to achieve the maximal degree of excellence in every possible world? This may or may not be true, but we need not settle the issue here. Note that, according to Plantinga, two other excellence-making properties, knowledge and power, have values for the greatest possible being in every possible world. That is to say, even in a world in which the greatest possible being is the only thing that exists, the greatest possible being will attain a level of excellence corresponding to knowledge of all necessary truths and will have certain powers. What I am claiming here is that

25 Recall my remarks at the end of section 3.2. The intrinsic maximum of a property is the same in every possible world. So, the maximum degree of excellence that Plantinga claims the greatest possible being has in every world is a constant finite positive value. I discuss this further in the next paragraph.
the greatest possible being is going to be as maximally excellent as possible in every world. Thus, I claim that there is a lower bound on \( h(B_i) \), where \( B_i \) is the greatest possible being, as follows:

For any natural number \( i \), there exists some positive real number \( k \) such that \( h(B_i) > k \) or \( h(B_i) = k \).

That is to say, for any natural number \( i \), each \( h(B_i) \) in the infinite series \( g(B) \) will be greater than or equal to some positive value of excellence \( k \). The same argument I gave for \( g(B) \) being unbounded above applies \( \textit{mutatis mutandis} \) to this new qualification of the greatest possible being’s excellence. We may conclude that the property of greatness (as Plantinga has defined it) does not exhibit an intrinsic maximum. Therefore, the property of maximal greatness is impossible; in other words, Plantinga’s first premise, (29) above, is false.

V. Conclusion

So, where does this leave Plantinga? It seems to me that he is caught between two conflicting requirements for a successful argument. These two are:

(a) Necessary existence must be a quality of the property of maximal greatness; and

(b) maximal greatness must have an intrinsic maximum to be consistent and coherent; that is, so that (2) above may be held as a premise.

If Plantinga fails to satisfy (a), then he is subject to his own criticism of Anselm’s arguments (see section 1.4). If he fails to satisfy (b), then his argument is subject to a version of his response to Gaunilo’s original objection (section 1.2 above). That is to say, failing to satisfy one gives only the conclusion that God exists in a maximally excellent way in \textit{some world or other}.\(^\text{26}\) But, by satisfying (a), Plantinga has made the property of “maximal greatness” conceivable only by a fool who deserves our sympathy.\(^\text{27}\)

\(^{26}\) See Section 1.4 of this paper, \textit{The Nature of Necessity} 202–5, and God, Freedom, and Evil 101–204.

\(^{27}\) I am indebted to Professors Jay Odenbaugh and Nicholas D. Smith at Lewis and Clark College. Also, I would like to thank Lara Sbordone for her help in proofreading the paper.
Works Cited


