The Realism in Moral Antirealism

Abstract

I investigate the implications of a certain puzzle which arises from metaethical antirealism. Here is the barebones outline of the puzzle: (i) if the usual antirealist theories are true, then any evaluative judgment *could* be "true" relative to certain people; (ii) but the judgment to disobey all moral judgments *could not* be true relative to anyone; (iii) therefore, the usual antirealist theories are not true. After spelling out the details, I show how antirealists may solve the puzzle if they are willing to take on board certain "realist" constraints. The goal of the paper is to further advance our understanding of the metaethical positions currently in play.

The Realism in Moral Antirealism

1. The Argument

Sharon Street (2006, p. 119) takes the defining claim of moral realism to be this: there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes.[1] On this definition, *non*-realists hold that all evaluative facts and truths, if there are any, depend upon (or describe) evaluative attitudes. Street thinks there are evaluative facts, such as the fact that parents should care for their children. So I will call her brand of non-realism 'factive antirealism'.[2] The task of this paper is to investigate a certain antirealism puzzle with an aim toward achieving a clearer vision of the line between realist and antirealist metaethical theories.

I'll give an initial synopsis of the puzzle and then consider its premises in more detail. We may state the puzzle as follows. Suppose that (i) moral realism is false and that (ii) there are moral truths or facts. Then a moral judgment *J* is "true" (or a fact) if and only if there are certain evaluative attitudes that *ground* or relevantly *explain J*'s being true. We may say that *J* is true *relative to* people who have the attitudes that explain why *J* is true. For ease of presentation, let us stipulate that *J* is true relative to people who tend to *endorse J*. So, for example, it is true (relative to *us*) that parents should care for their children because we tend to endorse the activity of parents caring for their children. Of course, the correct account of the dependence between moral judgments and evaluative attitudes might well be far more complex. But nothing in the argument turns on exactly *how* moral judgments depend upon attitudes.

[1] This claim is sometimes associated with moral *objectivism*.

[2] For the purposes of this paper, I make no assumptions about the nature of moral facts. Moral facts may be natural facts about non-representing mental states, for instance; or they may be *sui generis* moral truths. Even non-cognitivists, such as quasi-realists, may speak of moral facts (see Blackburn 1993).

Now it is logically possible for there to be people who endorse (or have attitudes that would explain the truth of) the following moral judgment: all moral codes ought to be disobeyed. Call that judgment 'Lawless'. If it is logically possible for people to endorse Lawless, then, given factive antirealism, it is logically possible for Lawless to be true relative to certain people. Now if Lawless is true relative to certain people, then according to Lawless, those very people ought to disobey Lawless itself: according to Lawless, people ought to disobey all moral codes, which include Lawless. But if those people ought to *disobey* Lawless, then it seems Lawless is not a moral code they ought to abide by (assuming that rules that ought to be disobeyed are not also rules that ought to be obeyed). Hence, Lawless is not a *true* evaluative judgment. Therefore, it is logically possible that Lawless is both true and not true relative to the same people. But no moral principle can be both true and not true relative to the very same people. Otherwise, one and the same set of evaluative attitudes could ground *contradictory* moral judgments.

The above contradiction arises from the starting hypothesis that a moral judgment *J* is true relative to a society if and only if the members of that society as a whole endorse (or have attitudes that would explain the truth of) *J*. Therefore, factive antirealism implies that a contradiction is logically possible. But contradictions are not logically possible. Therefore, factive antirealism is not true.

I have just sketched what I call 'the Self-Defeat Problem'. We may state the key premises as follows:

Premise 1: If factive antirealism is true, then possibly, Lawless is true relative to some people.

Premise 2: Necessarily, if Lawless is true relative to some people, then Lawless is false relative to those same people.

Therefore: If factive antirealism is true, then possibly, Lawless is both true and false relative to the same people.

Premise 3: Nothing can be both true and false relative to the same people.

Therefore: Factive antirealism is not true.

Let us review each of the premises and then turn to objections. Premise 1 is based upon the observation that the attitudes people form are *contingent* matters. Beings *can* value anything. We value honesty, charity, and truth, but not because rules of logic force us to do so. It is logically possible for there to be beings who despise honesty, charity, and truth. Indeed, for any judgment we make, we can imagine beings making the opposite judgment. More generally, for any evaluative attitudes and any judgment, we can imagine beings making that judgment on the basis of those attitudes. If this were not so, then there should be an explanation as to why not. What rules it out? In the absence of such an explanation, it seems that any judgment could be true relative to certain people, supposing that moral truth is grounded solely in certain attitudes. Therefore, if moral truth is grounded in certain attitudes, then it seems we can imagine *Lawless* being true relative to certain attitudes.

Consider, next, premise 2: necessarily, if Lawless is true relative to some people, then Lawless is false relative to those same people. According to Lawless, I ought to disobey all moral codes, including Lawless itself. In other words, anyone who obeys Lawless also *disobeys* Lawless. Hence, if Lawless were true relative to some people, then those people ought to both *obey* Lawless and *disobey* it. This is a contradiction if the following holds true: if a person ought to *disobey* a principle, then it is not the case that that person ought to *obey* that same principle. And even if that does not hold true, factive antirealism would still imply that there can be people who ought to both obey and disobey the same principle, and I suspect that hardly anyone would be happy with that result. (Alternatively, we could imagine a possible world where a community

endorses a pair of contradictory moral principles. Then, given factive antirealism, those principles, though contradictory, could both be true relative to the same people.[3])

We have seen that if factive antirealism is true, then Lawless *can* be true relative to some people: for it is possible for people to have morally relevant attitudes toward Lawless, such as the attitude of endorsing it. (That's possible even though it is impossible for people to actually *obey* Lawless.) Yet, it seems that nothing can be both true and false relative to the very same people (premise 3). Therefore, it seems we have a reason to doubt factive antirealism.

2. Exploring Solutions

I will now consider what I take to be the most penetrating and instructive proposed solutions I have encountered.

Solution 1: Lawless cannot be endorsed.

We saw that endorsing Lawless commits one to *not* endorsing it. This results in a contradiction, for something cannot both be endorsed and not endorsed. Therefore, it is not really possible to endorse Lawless, as Premise 1 asserts.

Reply. The phrase "commits one to not endorsing it" is ambiguous. The intended meaning is this: "entails that one does not endorse it". Otherwise, a contradiction is generated only if endorsing Lawless entails that one does not endorse it. But why think endorsing Lawless entails that one does not endorse it? It may be true that endorsing Lawless entails that one *ought not* endorse Lawless. But there is no contradiction in that: it is logically possible to endorse something without thereby abiding by the implications of that endorsement. (For example, I can endorse the value of not discriminating against people without even seeing all the specific implications of

[3] I owe this suggestion to [Removed].

that endorsement). So, it is not clear why we should think that simply endorsing Lawless is impossible. It hardly *seems* impossible.

We should distinguish here between abiding by Lawless, on the one hand, and endorsing it, on the other. No one can successfully abide by Lawless without at the same time not abiding by it (just by the meaning of 'Lawless'). But that is not to say that people cannot endorse Lawless. Endorsing and abiding are very different matters. I might endorse the value of honesty (especially when I catch someone being dishonest with *me*) without successfully following this code myself. If *endorsing* (or whatever attitudes ground moral truth on the version of antirealism we're considering) were the same as *abiding*, then no one could ever violate any true moral codes, because having attitudes that ground moral truths would then be sufficient for abiding by those truths. But we surely can violate true moral codes, if there are any. Therefore, endorsing and abiding are not the same.

Solution 2: Lawless is not a moral judgment.

Chris Gowans has suggested to me in correspondence that perhaps Lawless cannot be classified as a *moral* judgment. If Lawless is not a moral judgment, then there is no pressure to explain its truth in terms of attitudes. Lawless might be akin to a Liar sentence: it leads to paradox but only by a semantic trick, even if the exact nature of the trick is difficult to unpack.

Reply. Consider, first, that Lawless has certain earmarks of a moral judgment. For example, Lawless states what people *ought* to do: people *ought* to disobey all the moral codes. If Lawless does not count as moral, then it is not clear what the term "moral" is supposed to mean. One

wonders how we can decide in a principled way which propositions count as moral and which ones do not.

Moreover, there are *neighbors* of Lawless that appear to count as moral judgments without generating problems. For instance, <we should *obey* all moral laws> seems to be a true moral judgment. We might call it a "higher-order," since it's a judgment about how we should respond to moral judgments. There is no obvious problem with supposing that we *should* obey moral laws or with supposing that there are such higher-order moral judgments. But then if <we should *obey* all moral laws> may count as a (true) moral judgment, then why can't <we should *disobey* all moral laws> count as a (false) moral judgment? Consider other unproblematic neighbors: <we should disobey some moral laws>; <we should disobey all moral laws that we want do disobey>; <we should disobey all moral laws that we can't obey>. Each appears to be a genuine moral judgment. So, then, why shouldn't Lawless also count as moral? It seems to me that the burden is on the dissenter to supply a non-question begging explanation as to why certain judgments, such as Lawless, do not count as moral judgments.

None of these considerations are decisive. Perhaps someone will develop a clever account of moral judgments to explain why, for example, Lawless, but not certain of its neighbors, counts as a moral judgment. For example, one might try to motivate the claim no judgments about moral judgments are themselves moral. To be clear, it is far from clear *how* exactly the claim may be successfully motivated. After all, if endorsements give moral concepts their content, as antirealists may suppose, then once "first-order" moral concepts are in place, it seems perfectly possible to form judgments about our attitudes toward those concepts. For example, a community could surely fault a moral realist who never thinks carefully or charitably about moral antirealism—or, for that matter, about any first-order moral statements. "Politicians

should charitably consider whether outlawing abortion may do more harm than good" uses the term "good" and so is a second-order moral judgment, it seems. If this judgment isn't a *moral* judgment, we need an explanation as to why not. Such an explanation, if successful, would lead to significant progress in our understanding of moral judgments. What I suggest here, then, is that in the absence any such explanation, a theory that forbids us from treating Lawless as a moral proposition carries a cost—because judgments of the form *we should do X* are, in general, moral.

At this point someone might wonder whether the self-defeating nature of Lawless disqualifies it from being a genuine more principle. Perhaps Lawless is akin to a Liar sentence, such as "I am not now telling you the truth," which is true if and only if it isn't.[4] Consider that although Liar sentences may *appear* to express genuinely meaningful propositions, they do not: for if they did, then our universe would contain the contradiction that a certain proposition is both truth and not true. Perhaps Lawless is like that. Lawless *appears* to express a moral judgment, but its self-defeating nature reveals to us that it cannot in fact express a genuine moral judgment.

There is, however, an important difference between Lawless and Liar sentences. Liar sentences aren't *merely* self-defeating in that way that, say, "every sentence is meaningless" is self-defeating. Self-defeating sentences aren't by themselves paradoxical: they are simply false. Liar sentences are worse than false. In fact, they aren't *even* false, because if they were false, then they would also be true. Consider that the paradox of a Liar sentence arises solely from certain assumptions about language and logic. By contrast, Lawless, though self-defeating, isn't paradoxical—*unless* factive antirealism is true. But if the only reason to doubt that Lawless is

[4] For a recent discussion of Liar sentences and possible solutions to them, see [removed].

moral in nature is that it would otherwise cause trouble for factive antirealism, then we lack an *independent* reason to doubt that Lawless is a moral judgment.

Consider, furthermore, the dialectic. Suppose there is a philosophical debate over theory *T*. And suppose someone shows that, surprisingly, *T* entails that "every proposition is false" is possibly true. The entailment would be significant to see because "every proposition is false" is manifestly *not* possibly true. So, the entailment would call into question theory *T*—or at the very least, the entailment would expose a cost of *T*. But now suppose someone replies that the entailment only goes through if we assume that "every proposition is false" expresses a genuine proposition, and that "every proposition is false," like Liar sentences, might not express a genuine proposition. Should we then conclude that the alleged entailment poses no problem for *T*? It seems not: for "every proposition is false" seems to express a genuine proposition—one that can't be true. And unlike Liar sentences, no paradox arises from supposing that what *seems* true here *is* true. I suggest that Lawless poses a similar problem for factive antirealism. Lawless seems to express a genuine moral judgment—one that can't be true. Someone may deny this to protect factive antirealism from possibly entailing a falsehood. But without independent motivation for this denial, we lack a non-question-begging response to the Self-Defeat Problem.

It is worth emphasizing again that my primary task is to pursue a clearer understanding of the boundaries between realist and antirealist theories. I do not claim to show that no version of factive antirealism is sustainable. Perhaps there are factive antirealist theories that have the resources to motivate the claim that Lawless is not a moral judgment. I leave that open. In the next objection, I will discuss some *constraints* on developing a sufficiently nuanced version of factive antirealism.

Solution 3: You are knocking down a straw man.

Endorsing Lawless may be possible, but a robust antirealism will include some minimal consistency constraints for "true" moral judgments. We have no reason to think that Lawless would meet those constraints. Therefore, we have no reason to think that Lawless poses a problem for a sufficiently robust version of antirealism.

Reply

First, a concessive reply: if someone manages to devise a version of factive antirealism that entails that Lawless cannot be true, then I will grant that that version of antirealism survives the Self-Defeat Objection. The value of my argument would then be in shaving off versions of antirealism that are not sufficiently robust to survive the threat of self-defeat. The objection would help us develop a better antirealism.

However, it is not at all clear how to construct a factive antirealist theory that avoids self-defeat without that theory being intolerably *ad hoc*. Let us consider some attempts.

One attempt to rule out Lawless is to suppose that "true" moral statements must be rational to believe. However, that requirement fails to explain why Lawless cannot count as a "true" moral judgment. To see why, suppose, first, that epistemic rationality is solely a function of how things seem from one's "first-person" perspective. Then the problem is that we can imagine "brain-in-vat" scenarios in which something causes Lawless to *seem* true. For example, one's inner feelings about Lawless could be induced by a sophisticated neuro-stimulating machine. Or to use a more mundane example, imagine that someone is raised from birth to believe in Lawless; surely Lawless *could*, at some time, seem true to that person based upon everything that person has been aware of. So, it seems perfectly possible for "internalist" conditions on rationality to be met with respect to belief in Lawless.

Alternatively, suppose that rationality depends upon certain external conditions (such as being produced by a reliable belief forming process or a properly functioning cognitive system). The problem returns: it is perfectly possible for conditions of rationality to give rise to *false* beliefs, including a false belief in Lawless. Here is a scenario to draw this out. Imagine a world in which 99% of what you read on the Internet is true—a distant world, to be sure. You come across a webpage that says that experts have proven Lawless to be true by a subtle, complicated proof, and all objections have been shown to fail. You are five years old. And you believe the testimony. Surely, that belief could be rational for you. (It won't do to object that the belief is irrational on the grounds that the *specific* webpage, or section of the webpage, is unreliable, because that sort of objection, if successful, poses a problem for reliabilism itself—the so called "generality" problem[5]).

For a moment, I thought this: "Maybe we can escape the problem of self-defeat by relativizing the belief to an *ideal* rational agent, one that has knowledge of every possible factor relevant to assessing the moral proposition in question." We should be careful, however, to specify *our* relationship to the ideal observer. For if we suppose merely that moral judgments are true relative to certain ideal conditions that don't apply to us, then moral judgments won't be true relative to us. That isn't a favorable result, especially since factive antirealists think at least some moral judgments are true relative to *us*. If we suppose instead that a moral proposition endorsed by an ideal rational agent would be true for everyone (perhaps because the attitudes of each individual *would* necessarily converge under the relevant idealizations),[6] then we have essentially moved into the moral realist camp, since it is then true that at least some evaluative

[5] See Conee and Feldman 1998.

[6] Cf. Firth 1952 or Railton 2002.

facts or truths hold independently of all *our* (actual) evaluative attitudes—which is the defining thesis of the moral realism under discussion.[7] So, it is far from clear how conditions of rationality can save factive antirealism from self-defeat.

Someone might wonder if we should add *logical coherence* as a minimal requirement on a "true" evaluative judgment, for then we could rule out Lawless on the grounds that it is not logically coherent (because its truth implies its own falsehood). However, if "logical coherence" is a requirement on moral truth, then we have again strayed away from moral antirealism.[8] Consider that Lawless would be a necessarily false moral judgment *independently of anyone's attitudes*. (I will consider in a moment the objection that logical coherence itself depends upon people's attitudes.) Moreover, the judgment that we ought *not* to obey Lawless would be necessarily *true* independently of anyone's attitudes. Recall that the defining claim of moral realism, according to Street, is that there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes. The logical coherence requirement entails that very claim (if Lawless is indeed incoherent) and so is not part of the antirealism under investigation.

At this point, someone might wonder if it is consistent with the spirit of antirealist constructivist theories to suppose that certain "trivial" moral principles aren't grounded in anyone's attitudes. Let us say that a moral principle is "trivial" *iff* it trivially follows from true *metaethical* principles. Then perhaps an antirealist may suppose that Lawless is necessarily false on the grounds that it trivially follows from certain metaethical principles.

---

[7] It also follows that moral realism, as Street defines it, is compatible with certain versions of *constructivism*.

[8] Or, at least we have strayed from antirealism as characterized by Street (recall my opening paragraph). Certain expressivists, for example, may wish to include a coherence requirement while rejecting typical realist theories. Defining the boundaries between realism and antirealism remains an open inquiry.

[Removed] suggests how one might develop this "trivial truth" reply in his comments on the argument of this paper.[9] [Removed] writes,

> What can constructivists say by way of reply? I think that their best strategy is to affirm an ought-implies-can principle. The principle designed to make trouble for constructivists—Lawless, which says that one ought to disobey all moral codes—is, as [Removed] is well aware, impossible to obey. Without getting into any details at all about the best specification of an ought-implies-can principle, it seems plausible to argue that since we are incapable of adhering to Lawless, it is not the case that we ought to. And if it's not the case that we ought to adhere to it, then no contradiction can be derived from it, since the contradiction is supposed to take the form that the very same folks both ought and ought not to obey Lawless. So, it seems to me that the best way out of the problem that [author's name removed] sets for the constructivist is to deny that constructivism implies that we ought to obey Lawless. And constructivists can do this by means of endorsing an ought-implies-can principle.

In other words, a factive antirealist may suppose that the falsity of Lawless trivially follows from an ought-implies-can metaethical principle. Problem solved?

Let me first emphasize here that *if* [Removed]'s proposal succeeds, then we have learned something quite valuable: we have learned that in order to protect antirealism from the Self-Defeat Objection, the antirealist must include certain significant metaethical constraints. The ought-implies-can principle is not a trivial add-on. It is a deeply controversial principle which

[9] [Removed] for blind review.

many antirealists do not endorse.[10] So, if antirealists must in fact accept an ought-implies-can principle, or something like it, in order to keep antirealism from self-destructing, that would be valuable to know.

Unfortunately, however, there remains a problem even for antirealists who take on board an ought-implies-can principle. The problem arises from considering where the metaethical principles themselves come from. As [Removed] himself points out, one may ask where the metaethical principles derive their authority. [Removed] suggests a couple possibilities: (i) metaethical principles are ultimately derived from the endorsements of agents, or (ii) they are conceptual truths. Both options lead to difficulties, however.

Consider, first, option (i): metaethical principles are ultimately derived from the endorsements of agents. The problem here is that any principles that depend upon the endorsements of agents are merely *contingently* true. After all, agents could have made different endorsements. Thus, *if* metaethical principles have their authority on the basis of endorsements, then the metaethical principles aren't *necessarily* authoritative. The ought-implies-can principle, for instance, could have been false (or have been non-authoritative). The problem of self-defeat now re-emerges: possibly, people have attitudes that ground *both* the falsity of the ought-implies-can principle *and* the truth of Lawless. So, by the nature of the case, Lawless can be both true and false relative to the same people (since Lawless is true only if it is false). We have fallen back into the original contradiction.

Suppose we go instead with option (ii): metaethical principles are conceptual truths. We may still ask where the metaethical principles come from. What makes them conceptual truths? If our answer is in terms of *contingent* features of our minds (of how we think about things), then

[10] I am thinking, for instance, of antirealists who endorse *compatibilism*—that moral responsibility (and culpability for wrongdoing) is compatible with determinism.

the problem of self-defeat returns just the same: the ought-implies-can principle could have been false, and so Lawless could have been true (and false). To avoid falling into contradiction, it seems we must suppose that metaethical principles are necessary truths—true independently of what any of us thinks or feels or believes. But then the *ethical* truths that, per hypothesis, follow from the metaethical truths—*e.g.*, that one *ought not* endorse Lawless—also turn out to be necessarily true independently of what any of us thinks or feels or believes. In other words, there are first-order moral statements that *hold of necessity* regardless of anyone's attitudes. It seems we are now well inside the realist's camp.[11]

Antirealists may discover better strategies that are ultimately successful, and if they do, then we will have learned something valuable. For what it is worth, I have tried to find a plausible way to protect traditional antirealist views from the problem of self-defeat and have come up short. At this stage in the dialectic, it is simply not obvious how to save antirealism from self-defeat without building in distinctively *realist* metaethical constraints. Of course, there is this solution: build in the requirement that an evaluative judgment can only count as "true" if it does not entail the falsity of factive antirealism. But that way is patently *ad hoc*.

It seems to me, therefore, that the best hope for antirealists is to build in certain realist metaethical constraints. Consider, for instance, the following theory, which I call 'quasi-antirealism':

> Quasi-Antirealism: every evaluative fact depends upon certain attitudes, *except* evaluative facts that are strictly deducible from necessarily true metaethical principles.

Quasi-antirealists may escape the problem of self-defeat. They may escape because the fact that we ought not to obey Lawless is strictly deducible from what may be a necessarily true metaethical principle that, for example, ought-implies-can. Here is the upshot: if we are to avoid

[11] Even quasi-realists generally don't go this far: they typically say that such facts as "it is a mind-independent fact that stealing is wrong" obtain *because of our attitudes*—and could have failed to obtain.

the problem of self-defeat, then it seems we need a better, more nuanced version of antirealism—one that looks significantly more like realism.

3. *Tu Quoque*?

Aren't there scenarios in which moral *realism* implies that one ought and ought not to perform a particular action. Consider, for example, the following scenario:

> Your friend has fallen out of a boat and calls for a life ring. But unbeknownst to you, there are TNT explosives hidden inside the raft. They will explode if you toss the raft to your friend. What should you do?

The only sensible answer is that you both *should* toss the ring (given your state of knowledge) and should *not* (given the bad result). Right?

Notice, first, that this sort of problem does not discriminate between realism and antirealism; it targets all factive moral theories alike. So if it succeeds, it really is no help to factive antirealism.

Second, the problem posed here is importantly different from the problem of self-defeat. An easy response to the problem here is to make a familiar distinction between two senses of "should". One sense of "should" implies moral culpability. In that sense of should, you should surely toss the raft (given your beliefs). But there is another sense of "should" that has more to do with what you must to do to achieve a good end. In that sense, you should surely not toss the raft. This sort of reply seems acceptable, but it does not help against the Self-Defeat Objection to moral antirealism. It doesn't help because Lawless can be defined to require that people should,

in *every* sense of "should", disobey all moral codes. Therefore, the two problems are not parallel.[12]

The "*tu quoque*" worry is instructive, however, for it highlights the tenacity of the Self-Defeat Objection. In particular, it makes clear that the threat of self-defeat does not depend upon *moral* intuitions we may or may not have. The lesson of Lawless is derived, not from premises about how people ought to behave, but from premises about what behaviors are merely logically possible. Unfortunately, it is not clear how the usual factive antirealist theories can escape self-defeat given those premises. So, we have a new argument against antirealism to think about.[13]

[12] That is not to say that moral realism doesn't face problems of its own. My goal here is only to show that the type of objection I have raised against antirealism isn't easily converted into an objection to moral realism.

[13] Acknowledgements…

Works Cited

Blackburn, Simon (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Conee, Earl and Feldman, Richard (1998). "The Generality Problem for Reliabilism," *Philosophical Studies*, 89: 1–29.

Firth, Roderick (1952). "Ethical Absolutism and the Ideal Observer,' *Philosophy and Phenomenological Research* 12: 317–45.

Railton, Peter. (2003). *Facts, Values, and Norms*, Cambridge University Press.

Street, Sharon (2006). "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies*, 127:109–166.