An Ontological Proof of Moral Realism
Michael Huemer

While there is no general agreement on whether moral realism is true, there is general agreement on at least some of the objective moral obligations that we have *if* moral realism is true. Given that moral realism might be true, and given that we know some of the things we ought to do if it is true, we have a reason to do those things. Furthermore, this reason is itself an objective moral reason. Thus, we have at least one objective moral reason.

In the following sections, I elaborate on the preceding argument. I begin by clarifying the conceptions of moral realism, practical reasons, and probability used in the argument, after which I set out the argument in more detail, offering support for each premise. I then address a series of potential objections. Ultimately, we shall see that, given that moral realism *might* be true, it *is* true.

1. Conceptual Background

1.1. *Practical Reasons: First Person vs. Third Person*

Practical reasons are considerations that count for or against behaving in a particular way. To say that there is a reason to perform an action is not to say that the action would be reasonable all things considered; it is only to say that there is at least *some* consideration in favor of the act. This consideration might well be outweighed by other considerations. For instance, I have a reason to smoke crack cocaine (it would provide immediate pleasure), but I also have stronger reasons against smoking crack (health risks, legal risks, financial cost).

It will be helpful to distinguish two kinds of reasons (or perhaps two senses of "reason"). To illustrate, take the case of Thurston. Thurston is very thirsty. He believes, justifiedly, that the glass before him is full of potable water. This gives him a reason to drink from the glass. But, though Thurston has no reason to suspect this, the glass is actually full of (an oddly odorless) petrol. In some sense, therefore, there is a reason for Thurston *not* to drink from the glass. I shall refer to the first sort of reason, the sort of reason Thurston has for drinking from the glass, as a "first person" reason, the idea being that this sort of reason is available from the agent's point of view at the time of action. First person reasons determine what it is rational to do, or what it makes sense to do, from the agent's perspective, or, given what the agent is aware of at the time of decision-making. I shall refer to the other sort of reason, the kind of reason that there is for Thurston not to drink from the glass, as a "third person" reason, the idea being that these sorts of reasons are ascribed from a third person point of view.[1] Third person reasons obtain in virtue of the actual circumstances surrounding the agent, regardless of whether the agent knows of them or has any reason to believe they obtain.

Typically, if one knows that some third person reason to Φ exists, then one thereby has a first person reason to Φ. It is also possible to have a first person reason to Φ

when there is no third person reason to Φ. This would occur when one has misleading evidence that some third person reason to Φ exists. In the above example, Thurston has a first person reason to drink from the glass but no third person reason to do so.

In the following discussion, unless otherwise specified, I shall use "reason" in the sense of a first person (agent-accessible) reason.

1.2. A Version of Moral Realism

The version of moral realism with which I am here concerned is a doctrine concerning *moral reasons*. For present purposes, I take it that moral reasons are a species of practical reasons having at least the following two characteristics:

a. First, they are *non-selfish*. That is, they are reasons for action that do not derive from an action's favorable relation to the agent's own interests.[2] Practical reasons that derive from such relations are to be classed as *prudential*, rather than moral reasons.

  Now, it may seem that I am here begging the question against ethical egoism, which holds that morality is properly based on self-interest.[3] In fact, however, I am not here laying down any substantive assumption, but merely making a stipulation regarding the content of the thesis to be defended: in what follows, I aim to demonstrate the existence of a certain sort of non-selfish practical reason. The argument will thus *refute* ethical egoism, not merely assume its falsity.

b. Second, moral reasons are *categorical*. That is, they are reasons for action that do not derive from an action's favorable relation to the satisfaction of the agent's desires.[4]

  Again, it may seem that I am begging the question against Humean accounts of morality, which hold that morality is properly based on desires. In fact, however, I am laying down no substantive assumption, but merely stipulating the thesis to be defended: in what follows, I will demonstrate the existence of a certain sort of categorical practical reason.

Now, the ordinary notion of "morality" may well carry more content than these two stipulations account for.[5] But I single out these two conditions – non-selfishness and categoricalness – because they are controversial and highly theoretically significant features of the view of ethical reasons that I wish to defend. In fact, the sort of practical reason defended in section 2 will, in addition to satisfying these two conditions, be an intuitively clear example of a moral reason.

The version of moral realism in which I am interested holds that there are some objective moral reasons; that is, there are at least some objective facts of the form "S has a moral reason to Φ." What is objectivity? A fact is said to be objective when it obtains independent of the attitudes of observers – for instance, independent of whether observers believe it obtains, whether observers want it to obtain, whether observers value the fact, and so on.[6] But the terms "objective" and "subjective" have so many uses that, for the sake of clarity, I shall hereafter use the terms "observer-independent" and "observer-dependent." A reason for action will be observer-independent ("objective") in the relevant sense just in case the agent has that reason for action independent of observers' attitudes toward the agent and the kind of action in question.

To illustrate, consider the metaethical theory of cultural relativism, according to which the moral wrongness of an action simply consists in its being of a kind that is disapproved of in a particular society. Assuming that the relativist accepts that the

wrongness of an action constitutes a reason not to perform it, the relativist must hold, for example, that one has a reason to avoid murder simply in virtue of the fact that one's society disapproves of murder. This reason would be *observer-dependent* ("subjective") in the sense I have been discussing, because it depends upon the attitudes of observers toward the type of action in question.

Putting together the above points, my doctrine of moral realism holds that there are some practical reasons that are (a) non-selfish, (b) categorical, and (c) observer-independent.

1.3. Epistemic Probability

In the discussion to follow, some use is made of the notion of probability, and in particular, of the idea that some claim has a nonzero probability. This is to be understood in the sense of epistemic probability. The epistemic probability of a proposition is a measure of the degree of justification the proposition has in light of one's current evidence. Epistemic probability is thus in principle relative to an individual – a proposition may have different probabilities for different individuals, if those individuals have different evidence. Nevertheless, all parties to the debate about moral realism have sufficiently similar evidence that some generic observations about the probability of moral realism can be made – for instance, we may say generically that moral realism has a nonzero probability, provided no party to the debate has a conclusive refutation of moral realism.

A proposition has probability one, in the epistemic sense of "probability," if and only if we have conclusive justification (the strongest justification possible) for believing it. This applies to at most very few propositions, such as perhaps the proposition that something exists, or the proposition that 2 is less than 3. A proposition has probability zero for us if and only if we have conclusive justification for *denying* it. Thus, the proposition that I do not exist has (for me now) probability zero, as does the proposition that 1 equals 4. Generally speaking, the only propositions with probability zero are those that are contradictory or otherwise absurd.[7]

1.4. The Probability of Moral Realism

Moral realism, as defined, is a highly controversial philosophical doctrine. It is rejected by ethical egoists (though this hardly renders it very controversial). It is also rejected, as I have suggested, by cultural relativists and subjectivists. Moreover, it is rejected by Humeans – surely the dominant faction in modern debates about practical rationality – who hold that all reasons for action are desire-dependent, that is, that only an action's favorable relation to something the agent desires can give an agent a reason to perform that action.

But while the view is thus highly controversial, moral realism is not contradictory or absurd. Intelligent and rational philosophers have held the view, the most important of these being Immanuel Kant, and some thoughtful and informed participants in contemporary philosophical debate continue to endorse it. Almost no one, not even confirmed Humeans, relativists, or egoists, would say that this version of moral realism has been *conclusively refuted*. Moral realism thus has a non-extreme probability and is an interesting subject for debate.

2. A Proof of Moral Realism

2.1. The Probabilistic Reasons Principle

I begin with a principle that I call the Probabilistic Reasons Principle. The rough idea is that if some fact *would* (if you knew it) provide a reason for you to behave in a certain way, then your having some reason to believe that fact obtains also provides you with a reason to behave in the same way. Even a small epistemic probability of the fact's obtaining provides you with a (perhaps very small) first person reason for action.

Consider first an illustration involving prudential reasons. Anne is considering buying a particular lottery ticket. If she knew the ticket would win, that would be a prudential reason for Anne to buy the ticket. Therefore, the Probabilistic Reasons Principle tells us, if Anne merely has some reason to think the ticket will win, then she thereby has a first person reason to buy the ticket. There is in fact *some* reason to think that the ticket win, namely, that some ticket will win, and this one is as likely as any of the others (more simply: the ticket has a chance of winning). So Anne has a reason to buy the lottery ticket. Of course, this reason might be very weak, and it might be outweighed by the cost of the ticket. Nevertheless, I take it that the Probabilistic Reasons Principle gives us the intuitively correct verdict in this case. If tallying up the considerations for and against buying the ticket, Anne should not leave the "for" column blank; she should at least list something in the neighborhood of "the ticket might win" as a consideration in favor.

The Probabilistic Reasons Principle applies also to moral reasons. Suppose Jack likes firing his gun off in random directions in the woods. If he knew there was another person in the woods who would be hit by one of his bullets, this would constitute a moral reason not to fire the gun off. Therefore, pursuant to the Probabilistic Reasons Principle, if there is even a chance that a bullet would hit someone, Jack thereby has a moral reason not to fire the gun off. Of course, the lower the probability of hitting someone, the weaker this reason is. Nonetheless, as long as there is any chance at all of hitting someone, Jack has *at least some* reason to refrain.

That was a case of factual uncertainty, uncertainty about the non-moral circumstances. We can also give a case involving *moral* uncertainty. Suppose that Lisa is thinking of ordering lamb chops in a restaurant. But then she recalls that some people argue that eating meat is wrong. Lisa is unsure what to make of those arguments. If she knew eating meat was wrong, this would be a reason to refrain from ordering the lamb chops. Therefore, given that Lisa is unsure of whether eating meat is wrong, she has at least some reason to refrain from ordering the lamb chops.

The Probabilistic Reasons Principle even applies to epistemic reasons. Suppose that, if you knew John asserted P, this would provide evidence that P. Suppose also that in fact you have *some reason* to think that John asserted P (perhaps Sally said that John said that P). In that case, you have some reason to believe that P. This may of course be insufficient reason on which to found a justified belief, but it is nonetheless some reason.

Thus goes the intuitive motivation for the Probabilistic Reasons Principle. But the principle requires refinement. Imagine that you are standing outdoors, and you have become concerned about the possibility of a meteor strike. You are unable to take shelter or do anything else to reduce the probability of being struck by a meteor. All you can do is walk from one place outside to another, equally exposed spot that is equally likely to be hit. Intuitively, you have no reason to walk to another spot.

But suppose we argue as follows: if you knew the spot you are currently standing on

was about to be hit by a meteor, this would be a reason to relocate. Therefore, according to the Probabilistic Reasons Principle, if there is any chance that that spot will be hit by a meteor, you have a reason to relocate. And in fact, there is a nonzero chance that that spot will be hit by a meteor; hence, you have reason to walk somewhere else.

We could simply accept this conclusion, adding that your reason to relocate is exactly counter-balanced by a reason *not* to relocate (namely, the equal chance that you will be hit by a meteor as a result of your relocating). But it might be more natural to say that there is no reason either to move or not to move. If some location were less likely to be hit by a meteor than your current location, *then* you would have a reason to move to a safer place. But, we might say, in a case of exactly counter-balanced alleged reasons both for and against an action, the agent simply has no reason for or against.

Thus, we may wish to add a clause to the Probabilistic Reasons Principle, a clause that rules out cases of exactly counter-balanced (alleged) reasons. The following statement should suffice (though it may be stronger than necessary):

The Probabilistic Reasons Principle:
    If the following conditions hold –
        a.   If S knew that P, this would provide a reason for S to Φ,
        b.   If S knew that ~P, this would provide no reason for S *not* to Φ, and
        c.   S has some reason to believe that P
    – then S thereby has a reason to Φ.[8]

The meteor strike case is excluded because condition (b) is unsatisfied: if you knew your current location will *not* be hit by a meteor, this would provide a reason *not* to relocate, since you might be hit by a meteor as a result of relocating. So this formulation of the Probabilistic Reasons Principle does not require us to say that you have a reason to move somewhere else. (It doesn't *preclude* our saying that; it simply remains silent about the case, as it remains silent about all cases in which conditions (a)-(c) fail to be jointly satisfied.) The antecedent of the Probabilistic Reasons Principle postulates a stark asymmetry between P and ~P: roughly speaking, the truth of P would support Φing and the falsity of P wouldn't oppose it. In such a situation, if there's any chance at all that P is true, one has a reason to Φ.

What is the import of the qualifier "thereby," which appears in the consequent of the Probabilistic Reasons Principle? The qualifier is intended to specify a connection between the antecedent and the consequent – that is, when conditions (a), (b), and (c) hold, S has a reason to Φ *because of* or *in virtue of* that very fact.[9]

2.2. An Argument Against Recreational Baby Torture

I turn now to an argument against torturing babies for fun, which I shall refer to as the "Anti-Torture Argument." It is not an argument that it is *wrong* to thus torture babies (such a conclusion being too ambitious for one philosophy paper). It is only an argument that we have *some reason* to avoid torturing babies. This argument is also *not my argument for moral realism*; the argument for moral realism will appear in section 2.3 below.

The Anti-Torture Argument applies the Probabilistic Reasons Principle to the problem of baby torture, thus:

1.   If the following conditions hold –

       a.   If S knew that P, this would provide a reason for S to Φ,
       b.   If S knew that ~P, this would provide no reason for S *not* to Φ, and
       c.   S has some reason to believe that P
   – then S thereby has a reason to Φ.

2.   If we knew torturing babies was objectively wrong, this would provide a reason to avoid torturing babies.
3.   Even if we knew torturing babies was not objectively wrong, this would provide no reason *to* torture babies.
4.   We have some reason to believe that torturing babies is objectively wrong.
5.   Therefore, we have a reason to avoid torturing babies.

The reasons for accepting premise (1) have already been discussed. Premise (2) is true in virtue of the meaning of "objectively wrong." By definition, if something is known to be objectively wrong, then we thereby have a reason to avoid it. Of course, it is controversial whether anything is in fact objectively wrong. But there is no controversy as to whether an act's known objective wrongness would constitute a reason to avoid it.

Premise (3) is not analytic but nevertheless seems obviously correct. The failure of baby torture to be objectively wrong would not give us a reason *to* torture babies; it would merely *fail* to provide a reason *against* baby torture. If baby torture were not objectively wrong, this would be because some alternative to moral realism was correct – perhaps relativism, subjectivism, non-cognitivism, egoism, or even nihilism. But even if we knew one of those alternative theories was correct, none of them would give us a reason to torture babies.

But here is an objection to (3): Suppose Sadie the sadist knows that she would derive great pleasure from torturing babies, but she has so far refrained, solely because she thinks such torture would be objectively wrong. If in fact it would not be objectively wrong, then Sadie would have reason to torture babies. So there are some agents for whom (3) would be false.[10]

There are two replies to the objection. First, even for Sadie, the fact that baby torture was not objectively wrong would not itself provide a reason for torturing babies; instead, it would be Sadie's sadistic desires that would provide the reason. Second, it does not matter in any case if there are some agents to whom the Anti-Torture Argument does not apply. The goal of the present section is only to establish the existence of at least one reason, applicable to at least one agent, for avoiding baby torture. In the following section, we will see that this reason is an objective moral reason. Since moral realism maintains only that there are *some* objective moral reasons, it will suffice that there be at least one agent who has such reasons.

Why believe premise (4)? As discussed earlier (section 1.4), moral realism might be true. The theory is not absurd, nor has it been conclusively refuted. Furthermore, it is generally agreed that *if* moral realism is true, then baby torture would be one of the things that would be objectively wrong. So there is at least some reason to think baby torture is objectively wrong. Perhaps it is a weak reason, insufficient to justify the belief that baby torture is in fact objectively wrong. But for purposes of this argument, we only need some reason, any reason at all, for thinking baby torture to be objectively wrong.

Conclusion (5) follows from premises (1)-(4). Obviously, very few would disagree with (5), and most would consider it laughably weak. But the Anti-Torture Argument

will help us establish the much more controversial thesis of moral realism.

2.3. The Reason to Eschew Baby Torture Is an Objective Moral Reason

We have just seen that we have a reason not to torture babies. What sort of reason is this? Is it a selfish reason? Is it a desire-based reason? Is it an observer-dependent reason?

It is none of these things. The practical reason established by the Anti-Torture Argument is itself a non-selfish, categorical, observer-independent reason. There is a short argument for this, and a longer argument. I begin with the longer argument:

6. The premises of the Anti-Torture Argument are independent of interests, desires, and attitudes (in the sense relevant to moral realism).
7. The premises of the Anti-Torture Argument logically entail its conclusion.
8. If P is independent of interests, desires, and attitudes (in the relevant sense), and P entails C, then C is independent of interests, desires, and attitudes (in the relevant sense).
9. Therefore, the conclusion of the Anti-Torture Argument is independent of interests, desires, and attitudes (in the relevant sense).

Premise (6) is to be read as saying that each of the premises of the Anti-Torture Argument is true, and its truth does not depend upon any interest or desire of the agent, nor upon any attitude of observers toward baby torturers or acts of baby torture. This is the sense of "independent of interests, desires, and attitudes" that is relevant to establishing moral realism.

Why should we believe (6)? Consider the first premise of the Anti-Torture Argument, which is the Probabilistic Reasons Principle. The Probabilistic Reasons Principle is a general principle of rationality. Its status is similar to that of other axioms of rationality, such as the principle that rational preferences are transitive, or the principle that if one desires some end and one believes that a certain action will lead to that end, then one has a reason to perform that action. These principles appear to be necessary truths, true in every conceivable circumstance. (Even if, for example, one has no preferences, it is still true that *if* one prefers A over B and prefers B over C, then one is rationally committed to preferring A over C.) Of course, one's interests and desires may affect whether one in fact has a reason to Φ. But no matter what desires and interests one has – even if one somehow has no interests or desires – it remains true that *if* the knowledge of P would give one a reason to Φ, and the knowledge of ~P wouldn't give one a reason not to Φ, *then* a chance of P's being true gives one a reason to Φ. Nor does the truth of the Probabilistic Reasons Principle depend on anyone's attitudes toward baby torture – it is not as though, if we started approving of baby torture, then the Probabilistic Reasons Principle would somehow be falsified. So premise (1) is true independent of interests, desires, and attitudes in the relevant sense.

Premise (2) is also independent of interests, desires, and attitudes. Premise (2) states:

2. If we knew torturing babies was objectively wrong, this would provide a reason to avoid torturing babies.

This is an analytic truth and so is necessary in the strongest sense. That is, because (2) is

true in virtue of the meaning of "objectively wrong," its status is similar to that of "all bachelors are unmarried" and "triangles have three sides," which are true in every conceivable circumstance.

What about premise (3) –

3. Even if we knew torturing babies was not objectively wrong, this would provide no reason *to* torture babies.

? The reason this is true is simply that the failure of baby torture to be objectively wrong would be a mere absence of a reason to avoid torture, rather than a reason to torture. There is no desire or interest that we need to have for (3) to be true, nor is there any attitude that any observer needs to take towards acts of baby torture; if no one had any desires or interests and no one had any attitudes about baby torture, (3) would still be true.

Finally, consider premise (4) –

4. We have some reason to believe that torturing babies is objectively wrong.

This premise is true in virtue of the nonzero probability that moral realism is true, together with the likelihood that baby torture is objectively wrong *if* moral realism is true. For (4) to be the case, no one need have any interests or desires, nor need observers take any special attitudes. The truth of (4) thus is not dependent on interests, desires, or attitudes in the relevant sense.

Thus goes the motivation for premise (6). The rest of the argument is straightforward. (7), the claim that the premises of the Anti-Torture Argument entail its conclusion, is uncontroversial. Step (8) claims that if P is independent of interests, desires, and attitudes, and P entails C, then C is also independent of interests, desires, and attitudes. Why? Suppose that C *were* dependent on some interest, desire, or attitude. Then if that interest, desire, or attitude did not exist, C would be false. But then, since P entails C, whenever C is false P must be false. So if the interest, desire, or attitude did not exist, P would be false. So P is dependent on that same interest, desire, or attitude. Conversely, if P is independent of any interests, desires, or attitudes, then so is C.

Finally, we come to conclusion (9). None of the premises of the Anti-Torture Argument depends on any interest, desire, or attitude. So neither does the conclusion. Since the conclusion asserts that we have a certain reason for action, there is a reason for action that we have independent of interests, desires, or attitudes. That is, moral realism is true.

Now, here is the shorter argument to the same conclusion: The Anti-Torture Argument establishes that we have a reason for avoiding baby torture. What is this reason? It is *that baby torture might be objectively wrong.* But *that baby torture might be objectively wrong* is obviously not an appeal to some desire, interest, or observer attitude. Therefore, it can only be an objective moral reason.

3. Three Brief Objections

3.1. *The Argument Relies on Intuition*

Objection: The reason we think that if moral realism is true then torturing babies is objectively wrong is merely that we have the *intuition* that baby torture is wrong. The

Anti-Torture Argument therefore relies upon ethical intuition. But we have no proof that ethical intuition is reliable, and many have argued that it is unreliable.

Reply: If the Anti-Torture Argument claimed that baby torture *is wrong*, then we might need the assumption that ethical intuition is reliable. Likewise, if there were some information-source, or alleged information-source, that told us that we *should* torture babies, then we might need the assumption that ethical intuition is more reliable than that source, in order to avoid the situation in which our alleged reason against baby torture would be exactly counterbalanced by an equal reason in favor (following the suggestion of section 2.1 that exactly counterbalanced alleged reasons would cancel each other out). But in fact, we have *only* the intuition that baby torture is wrong, and no intuition, nor any other putative information source, supporting baby torture. So if intuition even *might* be reliable, then we have at least *some* reason to avoid baby torture since it *might* be wrong. And even the most hardened moral skeptics will find it difficult to maintain that there is zero probability that ethical intuition is reliable.

3.2. The Reason to Avoid Torture Is Non-Objective

Objection: Of course we have a reason not to torture babies. Subjectivists, cultural relativists, and Humeans all agree with this. They just think that our reason to avoid baby torture derives from conventions, emotions, or desires. All of the premises and the conclusion of the Anti-Torture Argument are consistent with this. So the Anti-Torture Argument can't show that moral realism is true.[11]

Reply: The Anti-Torture Argument of section 2.2 is not the argument for moral realism. The conclusion of the Anti-Torture Argument is merely that we have a reason to avoid baby torture, which is indeed compatible with most non-realist theories. The argument for moral realism consisted of steps (6)-(9), and the discussion surrounding them, in section 2.3. The present objection says nothing to address that argument.

Similarly, the objection that the reason for avoiding baby torture depends on one's having a *desire* to be moral falls flat, because it simply ignores the argument already given in section 2.3 that this is not the case. The reason for avoiding baby torture that we are considering is *that torturing babies might be objectively wrong*. This is not a desire-dependent reason. If an action is *objectively wrong*, this constitutes a desire-independent reason to avoid the act, because that is just part of what "objectively wrong" (in my usage) means. If an action merely *might* be objectively wrong, the fact that it might be objectively wrong also constitutes a desire-independent reason to avoid the act. The only way to avoid this would be to claim that the Probabilistic Reasons Principle is somehow dependent on desires for its truth, a suggestion already rejected above (section 2.3).

3.3. The Probabilistic Reasons Principle Is Non-Objective

Objection: The Probabilistic Reasons Principle, while perhaps correct in some sense, is not *objectively* correct. The Probabilistic Reasons Principle is an evaluative claim; it describes a condition under which one has a *reason* to behave in some way, and all such claims are evaluative. Furthermore, all evaluative claims are non-objective – perhaps they are dependent on individual values or other attitudes, or perhaps they are dependent on social conventions.

Reply: This represents a coherent way of avoiding the argument for moral realism, but it is not a very attractive way. There are two reasons why it is less plausible to deny objective truth to the Probabilistic Reasons Principle than it is to deny objective truth to

standard examples of moral judgments:

First, the Probabilistic Reasons Principle is not a claim about morality *per se*, like the claim that stealing is wrong or that everyone has a right to life. The Probabilistic Reasons Principle is a general, formal principle governing reasons of any kind, including prudential reasons, instrumental reasons, and even epistemic reasons, in addition to moral reasons. Thus, to deny objective truth to the Probabilistic Reasons Principle on the grounds that it is a claim about reasons would apparently involve one in anti-realism about reasons in general, which I take to be much less attractive than mere *ethical* anti-realism.

Second, the Probabilistic Reasons Principle is not a categorical reason-ascription. That is, it does not say that anyone in fact has a reason for anything. The Probabilistic Reasons Principle says that *if* certain reason-claims hold, *then* another, closely related reason-claim holds. And whether or not it is objectively true that anyone has a reason for anything, it is plausible that conditional claims of this sort could still be objectively true. For example, it is objectively true that *if* fetuses have a right to life then killing them is prima facie wrong, even if it is not objectively true that fetuses have a right to life. Likewise, it is plausible to hold that certain broad rationality constraints are objectively correct, even if there is no objective fact as to what particular reasons we have. For instance, plausibly, it is an objective fact that rational preferences are transitive, even if there is no unique set of preferences that rationality demands. Similarly, it is plausibly an objective fact that *if* the known truth of P would be a reason for Φing and its known falsity would provide no reason against Φing, *then* a chance of P's being true provides a reason for Φing – even if there is no objective fact as to whether the known truth of P actually provides a reason for Φing.

4. The Theory-of-Reasons Objection

I devote a separate section to one final objection, because I find this objection the most serious and interesting. The objection, which I shall call the "Theory-of-Reasons Objection," claims that the Probabilistic Reasons Principle is not universally true. Granted, the Probabilistic Reasons Principle applies in all the sorts of cases used to motivate the principle in section 2.1. It applies in cases of uncertainty about whether an action would cause some desired outcome, uncertainty about whether an action would cause some morally good outcome, uncertainty about whether some outcome that an action would cause counts as morally good, and even uncertainty about whether some piece of evidence for some belief exists. But the principle fails for the one class of cases where the argument of this paper needs it: the case of uncertainty *about the theory of reasons*.

At first glance, this may seem an ad hoc suggestion, no better than a bare denial of the first premise of the argument. To show why this is an interesting objection, I will spend sections 4.1-4.3 motivating and refining the Theory-of-Reasons Objection. I will then respond to it in section 4.4.

4.1. Rationality Facts Do Not Inherently Provide Reasons

To begin with, consider how "rationality facts" – specifically, facts of the form "It is rational to Φ" or "S has a reason to Φ" – relate to reasons for action. Do these facts themselves constitute reasons for action? Suppose that I am deliberating about whether to Φ. I list the reasons for and against Φing. After rationally weighing these considerations, I come to the conclusion that overall, I have most reason to Φ, so it would be rational for me to Φ. Have I just now discovered *another* reason to Φ, namely, that Φing would be rational?[12] There are three reasons to think not.

To begin with, there is the suspicion that to count the rationality of Φing as a reason for Φing is simply a confusion. To say that Φing is rational is just to say that the balance of reasons favors Φing. It is to describe some overall feature of the existing reasons. To think of this as itself reporting one of the reasons seems confused, in approximately the way that it would be confused to think of a bank account balance as a revenue source.

Second, and more seriously, it seems as though, if we treat rationality facts as providing reasons for action, there will be too many reasons for action. Thus, suppose that I am deciding between two courses of action, Φ and Ψ. In my initial deliberations, I find that the reasons favoring Φ *just barely* edge out those favoring Ψ. But then I reflect: there is another reason favoring Φ that I forgot to list, namely, that Φing is overall most rational. In fact, this new reason for Φing is a rationally conclusive reason. So actually Φ *decisively* beats Ψ, in the full weighing of reasons.

Third, if we treat facts about reasons as themselves providing reasons, then we swiftly generate an infinite regress of reasons. Suppose A is a reason to Φ. Then the fact that there is this reason to Φ is itself a second reason to Φ. But then, the fact that this second reason to Φ exists is a third reason to Φ. And so on.

4.2. Applying the Probabilistic Reasons Principle to Theoretical Uncertainty about Reasons

So let us assume that rationality facts do not provide reasons for action. Now consider the following case. Some theory of practical reasons holds that the fact that P is a reason to Φ. But the theory is false, and the fact that P is not actually a reason to Φ. But, though the theory is false, agent S is not absolutely certain that it is false. For S, the epistemic probability of the theory is 10%. On the other hand, S knows for certain that (whether or not the theory in question is true) there is *no* reason *against* Φing. It seems that all of this is possible. The question is, does S have a reason for Φing?

We might be tempted to answer yes. But suppose that S were to Φ. What rational motive could S have for doing so? Insofar as S is rational, S would not be motivated by the truth of P itself, since that does not constitute a reason to Φ. S would be moved instead by the thought that S might have a reason for Φing. Since S knows that he does not have a reason against Φing, he knows that he is not rationally required *not* to Φ, and he *might* be rationally required to Φ. So, to be on the safe side, S decides to Φ. But now it looks very much as though we are ascribing to S a goal of "being rational" or "doing what one has most reason to do."[13] That is, it looks as though some such goal is the only one in terms of which we could explain why a rational person would Φ in this situation.

But this is to ascribe to facts about rationality precisely the role that we denied to them in section 4.1. There, we said that in general, the fact that it would be rational to Φ, or that one has a reason to Φ, does not itself constitute a reason to Φ. If rationality facts of this form do not constitute reasons to Φ even when we know them to obtain, then a mere *chance* that it would be rational to Φ, or that one has a reason to Φ, does not

constitute a reason to Φ. So in the scenario we've been discussing, S would in fact have *no* reason to Φ.

So there is an exception to the Probabilistic Reasons Principle, specifically for cases in which a false theory of reasons has a nonzero epistemic probability. This exception blocks the argument to moral realism, for moral realism is a theory about practical reasons. If moral realism is *in fact* false, then we have no reason to do the things that the theory suggests we have reason to do (again, except insofar as they happen to correspond with whatever the correct theory recommends), even if, on our evidence, there is a nonzero epistemic probability of moral realism's being true.

4.3. Rationality Facts May Provide Desire-Dependent Reasons

Thus goes the Theory-of-Reasons Objection. But it seems that the general ban on treating rationality facts as providing reasons for action is too strong. Suppose, as might well be the case, that I have a basic desire to be rational. Maybe the idea of being a rational person makes me happy. Surely at least in this case, the fact that Φing would be rational would give me a reason to Φ.

The Theory-of-Reasons Objection should be revised to allow for this. Rather than claiming that rationality facts do not provide reasons for action, the revised Theory-of-Reasons Objection will claim that rationality facts *need not* provide reasons for action, and that if and when they do so, these reasons are dependent on one's particular interests or desires. If one has a desire to be rational or an interest in being rational, then the fact that Φing would be rational gives one a reason to Φ (additional to whatever reasons made Φing rational to begin with). Otherwise, it does not.

The Theory-of-Reasons Objection, thus revised, still poses a problem for my argument for moral realism. The problem now is that although the Probabilistic Reasons Principle *may* apply in cases of uncertainty about the theory of reasons, whether it does so will depend upon our interests or desires. As a result, the conclusion of the Anti-Torture Argument will not be independent of our interests and desires.

4.4. A Reply: Rationality Facts Provide Categorical Reasons

My response to the Theory-of-Reasons Objection will be that, contrary to the arguments of section 4.1, facts of the form "It is rational to Φ" and "S has reason to Φ" *do* provide reasons for action, and they do so categorically, for all rational agents, not merely because of some special desire that one might happen to have.

Let us start with a related case. God appears to you and, after proving His identity, presents you with a box with a small red button on it. God informs you that pushing the mysterious button would be very much in your interests, though He refuses to tell you anything about how it would promote your interests, or which interest or interests it would promote, and you are entirely unable to guess these things yourself. Nevertheless, you trust God's word. Would it be prudent to push the button?

It seems to me that the answer is obviously yes. But consider: what prudential motive could you have for pushing the button? None of your particular interests could be cited, because you don't believe, for any particular interest of yours, that pushing the button would promote *that* interest. All you have is the general belief that it would somehow promote your overall welfare. So your motive would have to be that of promoting your overall well-being, or: doing what you have the most overall, prudential (third person) reason to do. Furthermore, the case seems to show that *prudence requires*

that you have this generic motive, and not merely motives to promote each of your particular interests (a desire to promote your health, another desire to promote your own enjoyment, another desire to preserve your friendships, and so on).

Before moving on, I need to address an objection to this argument. One might say: You need not have a general motivation to promote your overall interests; particular motives to promote specific interests would suffice. When God tells you that pushing the button would promote your interests in some way or other, this information *increases the probability*, for each specific interest, that that interest would be promoted by pushing the button. Thus, you can derive a little bit of motivation to push the button from each of the particular interests that you now think the act might promote.

In response, we can modify the scenario as follows. God informs you that you have many interests that you do not currently recognize. It turns out, He says, that pushing the button would promote your overall interests solely by promoting some of *those* interests, that is, the ones that you are not currently aware of. God still refuses to tell you what these interests are, but He assures you that you will definitely be better off as a result of pushing the button. Again, you trust His word. Would it be prudent to push the button?

In this case, if you push the button, we cannot say that your motive for doing so would derive from your particular motivations to promote specific interests, since by hypothesis, you know that the specific interests that you currently recognize would not be promoted. So your pushing the button would have to be motivated by a generic desire to promote your overall interests. Again, it seems to me that it would clearly be prudentially rational to push the button; therefore, having this generic desire to promote your overall interests must be part of prudential rationality.

Consider now the analogous point concerning rationality in general. Suppose that God presents you with a box with a small blue button on it. This time, God informs you – and again, you completely trust His word – that from a third person standpoint, you have most reason overall to push the button. He refuses to tell you anything about what the third person reasons for pushing the button are, except that they are reasons you do not currently recognize. In this case, it seems to me, it would be rational to push the button. But your motivation for doing so would have to consist in a generic motive to do what you have the most third person reason to do. So having that motive must itself be part of rationality.

If this is correct, then the rational goal of doing what one has most third person reason to do can be used to support the argument for moral realism. It is this motive that would lead a rational person to eschew baby torture, given no reasons in favor of baby torture and a nonzero epistemic probability that baby torture is objectively wrong. Since this motive is part of rationality itself, the rationality of eschewing baby torture is independent of whatever desires and interests one happens to have.

4.5. Double-Counting and Infinite Regresses

What can be said in response to the arguments of section 4.1? There, we worried that to count rationality facts as reasons for action will result in there being too much reason to perform a given action – e.g., that in a case in which intuitively, one would say there was only slightly more reason to Φ than to Ψ, it will turn out that there is actually *much* more reason to Φ.

The response to this worry is simple. We need to distinguish the claim that P

provides *a reason* for action from the claim that P provides an *independent* reason for action. To illustrate the distinction, suppose I have an end, E. Suppose I know that Φing would promote M, which is desirable solely as a means to promoting E. Then I have the following instrumental reasons for Φing:

> That Φing would promote E.
> That Φing would promote M.

Each of these facts is clearly a reason to Φ, and they are clearly distinct facts. So I have two reasons for Φing. But they are not two *independent* reasons; rather, the second reason is fully dependent on the first; [Φing would promote M] counts as a reason to Φ only because [Φing would promote E] is a reason to Φ, and M will promote E. Because the two reasons are related in this way, we do not add the two reasons together in computing how much reason there is, overall, to Φ. Rather, we need only take the strength of the first reason.

Similarly, suppose A is a reason to Φ. Then there will be the following series of reasons to Φ:

> A.
> A is a reason to Φ.
> [A is a reason to Φ] is a reason to Φ.
> etc.

But these are not *independent* reasons to Φ. All of the reasons after the first one are fully dependent on the first reason. So in computing how much reason there is, overall, to Φ, we do not add these reasons together. We need only take the strength of the first reason.

Admittedly, the view I have just articulated postulates an infinite series. Whenever there is a reason to Φ, there is an infinite series of reasons to Φ (all but one of which are fully dependent on the first reason). However, not all infinite series are vicious; some infinite series are benign. For example, there is the famously benign truth regress: if any proposition, P, is true, then there will be the following true propositions:

> P.
> P is true.
> [P is true] is true.
> etc.

Almost no one regards this as a vicious regress; almost no one says that we must avoid this regress by denying that P entails [P is true].

Now, the infinite regress of practical reasons bears considerable resemblance to the truth regress. Both are infinite series of propositions. In both cases, each succeeding step in the series is formed by applying a certain predicate (the same predicate in each step) to the proposition that constitutes the previous member of the series. In both cases, each succeeding proposition in the series is held to be a consequence of, rather than a precondition on, the truth of the preceding proposition in the series. In neither case is there any actual physical or psychological process that must be infinite. Barring

arguments to the contrary, then, it appears that the practical-reasons regress is as benign as the truth regress.[14]

5. Conclusion

The argument of this paper shows that there are first person, non-selfish, categorical, observer-independent reasons for action. This results merely from the fact that we have some epistemic reason to believe that certain actions are objectively right or wrong. But the argument does not show that these practical reasons are particularly strong ones – if the epistemic probability of there being objective wrongness is very low, then the sort of practical reason established by my argument would be a relatively weak reason. Furthermore, nothing in this paper shows that there are any *third* person, non-selfish, categorical, observer-independent reasons for action. Some might therefore consider the form of moral realism we have established to be disappointingly weak.

I would therefore like to make some concluding remarks about the significance of the argument of this paper. The chief import of the argument is not immediately practical. It is not, for example, that as a result of the discovery of this argument, we will torture fewer babies. Our behavior may not change in any noticeable way. The significance of the argument lies rather in its implications for the general theory of practical reasons. The leading theory of practical reasons – or rather, the leading family of theories – is Humean. Humean theories hold that all reasons for action are desire-dependent in some fairly strong sense – they depend, if not on one's current actual desires, then on one's future desires, or the desires one would have after reflection, or the desires that one would desire to have, or something of this sort. Crucially, no Humean theory countenances purely *intellectual* sources of *practical* reasons. Humeans deny, that is, that a reason for action can be generated purely from beliefs. They also deny, incidentally, that anyone can possibly be *motivated* to do anything purely by beliefs. The Humeans apply their view not merely to third person reasons but to first person reasons. And they do not merely hold that beliefs cannot provide *very strong* reasons for action. Humeans hold that beliefs by themselves provide *no reasons whatsoever* for doing anything. Beliefs just are not in that line of work at all.

Furthermore, this broadly Humean view of practical reasons forms the main premise in what is perhaps the single most often repeated argument in all of the meta-ethics literature: roughly, the argument that there cannot be objective values because if there were such things, mere beliefs about them would suffice to provide reasons for action, and beliefs alone can never provide reasons for action.[15]

Pace the Humeans, the argument of this paper shows that an action can be motivated and justified, at least to some degree, solely on the basis of certain cognitive states. One can be rationally moved to do A because of the nonzero epistemic probability of a certain proposition. No remotely Humean theory can accept this. The argument, then, entails a rejection of the leading modern theory of reasons and rationality, and therewith the most common motivation for the leading meta-ethical theories of the past century.

One might wonder why I do not rely upon the following, simpler argument: Certainly there are people who *believe*, whether rightly or wrongly, that certain acts are objectively wrong. That suffices to provide *those* people with a reason to avoid those acts. Therefore, beliefs by themselves can provide reasons for action, contrary to the Humean theory.

The Humean might respond to this by saying that the belief in objective values is unjustified or irrational, and that it therefore does not provide genuine normative reasons for action – that is, it has no tendency to render actions motivated by it rational. But the parallel response could not be made to the *probabilistic* argument I have made in this paper, because it cannot seriously be maintained that it is unjustified or irrational merely to assign a *nonzero probability* to the existence of objective values. *Pace* David Hume, then, reason *can* be an influencing motive of the will.[16],[17]

---

[1]First person reasons are sometimes called "subjective" reasons, and third person reasons "objective." However, because the terms "subjective" and "objective" have so many uses (including a different use later in this paper), I have adopted the "first person"/"third person" terminology to avoid confusion.
[2]For this view of morality, see Kurt Baier, *The Moral Point of View* (Ithaca, N.Y.: Cornell University Press, 1958), 189-191, 308-15; William Frankena, *Ethics* (Englewood Cliffs: Prentice-Hall, 1963), 18; Steven M. Cahn, *Puzzles and Perplexities* (Lanham, Md.: Rowman and Littlefield, 2002), 68-9.

What is meant by an action's having a "favorable relation" to one's interests? Roughly, that the action would make one better off, or one has justification for believing that it would do so, or one's expected welfare conditional on one's performance of the action is greater than one's unconditional expected welfare, . . . or something in this neighborhood. I leave the phrase "favorable relation" in the text to avoid unnecessary debates about exactly what the relationship between an action and one's interests must be for one to have prudential reason to perform the action.
[3]For this view of ethics, see Thomas Hobbes, *Leviathan* (Cambridge: Cambridge University Press, 1996); Ayn Rand, *The Virtue of Selfishness* (New York: Signet, 1964). Plato can also be read this way in *The Republic* (Indianapolis: Hackett, 1974).
[4]For this view of moral reasons, see Immanuel Kant, *Foundations of the Metaphysics of Morals*, tr. Lewis White Beck (Indianapolis, Ind.: Bobbs-Merrill, 1959), 31-4; Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001), 42-5. On the notion of the "favorable relation" to desire-satisfaction, see note 2 above.
[5]On this extra content of "morality," see my "Values and Morals: Outline of a Skeptical Realism," *Philosophical Issues* 19 (2009): 113-30.
[6]For a more precise account of objectivity, see my *Ethical Intuitionism* (New York: Palgrave Macmillan, 2005), 2-4.
[7]An exception is the case in which a non-absurd proposition is one of an infinite class of similar alternatives. For instance, if one knows only that a given variable must take on exactly one value in a certain continuous range, then, in the standard treatment, each of the continuum many possibilities will be assigned probability zero. However, these cases are not relevant to the argument of this paper.
[8]The Probabilistic Reasons Principle is an indicative conditional, but the nested conditionals (a) and (b) are subjunctive.
[9]The need for the "thereby" qualifier is created by the tendency of philosophers to misinterpret "if-then" statements as material conditionals, that is, to interpret "If P then Q" as meaning "Either P is false or Q is true," such that there is no requirement that P and Q should have any connection to each other. My use of "thereby" blocks this interpretation.
[10]I thank Loren Lomasky for pointing this out, using a different example.
[11]I thank Justin Weinberg for something vaguely like this objection.
[12]I thank Sarah Raskoff for discussion of this issue. See also Niko Kolodny, "Why Be Rational?", *Mind* 114 (2005): 509-63.
[13]The sentence "S has the goal of doing what S has most reason to do" is ambiguous. One reading, which I call the *de dicto* reading (after a distinction in the philosophy of language literature), is the following: "S wants the following to be the case: [for all x, if S has most reason to do x, then S does x]." The other reading, which I call the *de re* reading, is this: "For some x, S has most reason to do x, and S wants the following to be the case: [S does x]." The *de dicto* reading is the correct one.
[14]For discussion of virtuous and benign infinite series, see my "Virtue and Vice Among the Infinite" in *Ad Infinitum: New Essays on Epistemological Infinitism*, ed. John Turri and Peter Klein (Oxford University Press, forthcoming). Note that the practical-reasons regress is clearly benign on the account defended in that

article, since the infinite series of practical reasons does not require anything to instantiate an infinite intensive magnitude.

[15]See David Hume, *A Treatise of Human Nature* (Buffalo, N.Y.: Prometheus), 413-18, 455-70; Michael Smith, *The Moral Problem* (Oxford: Basil Blackwell, 1994), 12; J.L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977), 40-41; Alan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, Mass.: Harvard University Press), 33; R.M. Hare, *Objective Prescriptions and Other Essays* (Oxford: Clarendon, 1999), 1; P.H. Nowell-Smith, *Ethics* (Baltimore, Md.: Penguin, 1954), 39-42; Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001); and Richard Fumerton, *Reason and Morality: A Defense of the Egocentric Perspective* (Ithaca, N.Y.: Cornell University Press, 1990), chap. 3.

[16]See Hume, *Treatise of Human Nature*, II.III.iii.