

A Defense of Categorical Reasons
Russ Shafer-Landau
University of Wisconsin, Madison
shaferlandau@wisc.edu

Draft version – do not cite.

I

Categorical reasons, as I will define them here, are reasons that obtain independently of their relation to an agent's commitments. Such reasons do not depend for their existence on their being instrumental to the achievement of any of an agent's desires, goals or cares. I believe that there are categorical reasons for action. I will offer two arguments on their behalf. I do not think that any of their premises can be plausibly denied.

If there are categorical reasons for action, then practical instrumentalism is false.

Practical instrumentalism (hence, just *instrumentalism*) is the view that the only reasons there can be are so-called *hypothetical reasons*, i.e., reasons to do things that are in some way ancillary to the achievement of one's commitments (cares, desires, wants, goals, etc.).

Apart from the intrinsic interest of the matter, showing that there are categorical practical reasons, and that instrumentalism is false, is important for at least two reasons. First, it would enable us to resist relativistic arguments that assume that moral requirements entail excellent reasons for action, but make reasons contingent on our commitments, thereby making the content of moral requirements contingent on our commitments. Second, it would provide us an adequate reply to arguments that assume a commitment-independent source of moral requirements, and then proceed, with the help of instrumentalism, to the conclusion that there may be no good reason to abide by morality's demands.

II

Both of my arguments for categorical reasons, and against instrumentalism, begin by directing our attention to a familiar sort of example: that of the dedicated, successful immoralist. Imagine a person who is very sharp, very cunning, but also deeply malicious. His happiness is directly proportioned to the misery he wreaks. His top priority in life is to cause pain and suffering, even if, as he knows, such conduct will likely bring an early death, or a long incarceration. We intuitively regard such a person as (at the least) morally obligated to desist from the cruel treatment he longs to impose. Don't we also believe that there are excellent reasons for him to so refrain—namely, all of those considerations that constitute the wrongness of his actions? The reasons to refrain from cruelty are (at the least) the very considerations that make his actions wrong in the first place.

Consider an experienced torturer working on behalf of an authoritarian government. Such a person not only endorses the legitimacy of the regime, but takes active pleasure in breaking his victims. His greatest joy is stripping the last vestiges of dignity from those who initially resist his demands. At a given session, as he is about to apply the electrodes, he pauses to consider the merits of his action. He sees that doing so will get him what he most wants, and will frustrate none of his desires. He proceeds accordingly.

Consider a different case, one in which a person can very easily rescue another. A child has strayed from her parents on a busy city street, and is about to toddle into the path of a car that will surely kill her. The bystander sees what is happening. He need only reach an arm down to the child to save her from an awful death. Rather than doing so, he watches in delight as the

child is killed by an oncoming car.

If there were nothing to be said against these actions and omissions—no considerations that opposed, extinguished or overturned the case these agents might make for their cruel conduct—then it is hard to see how their actions could be wrong. But they obviously are wrong. And the sorts of considerations just mentioned—those directly relevant to matters of justification (and, in this particular context, those that indict the agent’s cruelty)—are precisely what reasons are. Reasons are, by definition, considerations that favor or oppose, that make something appropriate, legitimate, or justified (or the reverse). So, if we think that there is a plausible story to tell about why the dedicated evildoer is wrong for indulging his inclinations, then we are committed to there being reasons for him to refrain. And this despite the fact that, by hypothesis, he’s got no commitments that would be furthered by his doing so.

But surely, some will say, the moral monster has *some* commitments that will be furthered were he to refrain from cruelty. And that shows that he will, after all, have some reason to refrain. But we’ve no grounds for thinking this a categorical reason. Whatever reasons he has to refrain will stem straightforwardly from his aversion to jail, or his desire to avoid the potential harms inflicted by his vengeful victims, etc.

There are two things we can say here. First, and fairly obviously, even if all real people in the real world do have at least some ends that would be served by avoiding cruelty, we can imagine a possible world in which our misanthrope does not. The instrumentalist’s rejection of categorical reasons is meant to express a necessary truth—reasons must further what an agent cares about—and so is vulnerable if there are possible contexts in which this truth fails to obtain. In the scenario I am imagining, the ruthless immoralist has no commitments that would be

served were he to refrain from his cruelty. But there are, nevertheless, excellent reasons for him to so refrain—namely, and at the very least, all of the considerations that make his proposed actions immoral.

More importantly, we don't want to make the case against cruelty dependent on an instrumental link with this man's goals. Suppose, for instance, that our torturer wanted to avoid the censure that he would earn were his actions publicized. The best way to minimize his risk is to stop doing what he does. Though this, let us grant, does provide him a reason to stop, it isn't the only, or nearly the strongest, reason to do so. The cruelties he perpetrates are opposed by a host of considerations that make no mention of his aims. These considerations are reasons—reasons to refrain from deliberately inflicting misery. And these reasons will, first and foremost, mention the suffering of his victims, and the absence of their consent to his treatment. If the immoralist's aversion to being found out enters into it at all, it is only in a subordinate role, as a consideration that may supply an additional reason to refrain from his actions, one that is likelier than the others to motivate him to do the right thing.

Here is the argument in a nutshell:

1. A reason = df. a consideration that favors, justifies, legitimizes or makes appropriate (or the reverse).
2. Some such considerations bear no instrumental link to an agent's ends.
3. Therefore there are some reasons that bear no instrumental link to an agent's ends.
4. Therefore there are categorical reasons.

I hope that the definition expressed in premise (1), though obviously not a reductive one, is nevertheless uncontroversial. It's meant to be broad enough to be acceptable both to fans and to critics of categorical reasons. Where they will part company is with regard to premise (2). I have tried to reveal its attractions with the example of the dedicated evildoer. So long as we think—as

all of us do—that there are genuine considerations to oppose his cruelty, and also think that such considerations obtain independently of his commitments, then premise (2) is secure. That would be enough to establish the existence of categorical reasons.

Instrumentalists will likely charge that premise (2) begs the question. It is true that the considerations offered in support of the premise, together with the definition expressed in (1), entail the falsity of practical instrumentalism. But that cannot be a fatal flaw. Rather, it is a requirement of any logically valid argument whose conclusion vindicates the existence of categorical reasons.

Still, the instrumentalist may claim that the examples and considerations that support premise (2) are insufficiently independent of the conclusion being argued for. An ideal argument is one that is not only logically valid, but one whose premises can find support from those who are as yet uncommitted on the matter at hand. If the only reason to endorse a premise is that one already accepts the conclusion it is meant to support, then the premise is question-begging. The instrumentalist will likely insist that the only ones willing to ratify premise (2) are those who are already committed to rejecting instrumentalism.

Now I don't think that instrumentalists are right about that. But before saying why, we might undertake a brief excursus on the matter of begging questions. Begging a question is sometimes unavoidable. In ethics, the likeliest scenarios are ones in which one is advancing fundamental normative or evaluative commitments. It is hard to avoid begging a question if one encounters someone who denies that pain is ever bad, or denies that there is anything immoral about humiliating vulnerable innocents. Perhaps the only way to avoid a *petitio* in these circumstances is to show that one's interlocutor is contradicting himself. This is the fond hope of

Kantians and others—to show that those with patently immoral sensibilities are in some way undercutting their own commitments and displaying some internal incoherence.

Perhaps the Kantians are right. It would be lovely were it so. But let us pursue other possibilities, ones that do not vindicate the existence of categorical reasons by attributing a contradiction to those who refuse to acknowledge them. On this alternative line, those who oppose our basic normative and evaluative commitments can coherently reject the claims we hold so dear. Any defense of our deepest commitments will have to come from the sorts of bolstering considerations that are involved in revealing a belief to be situated within a network of mutually supporting beliefs. But such a defense will not be able to avoid the charge of begging the question. The other beliefs we enlist on behalf of our original claim may be no more persuasive to opponents than the position originally in need of support.

Unless we can reveal a contradiction in our opponent's position, we may have to beg a question somewhere. The most likely point is, as I have said, with regard to our fundamental normative and evaluative commitments. And we are certainly in the neighborhood, when considering an endorsement (or rejection) of categorical reasons.

This is not yet to concede that this first argument is question-begging. But what if it were? There is independent reason for thinking that question-begging claims and arguments are ones that agents may sometimes be justified in believing. I am justified in believing myself to be conscious, even if others regard me as an unthinking automaton whose protestations are merely programmed behaviors. If I am imprisoned on false charges, I am justified in believing myself innocent, even if all publicly available evidence convinces everyone else of the justice of the sentence. If, having cried “wolf!” once too many times, my next cry goes unheeded, I am

nonetheless justified in believing that there is such an animal before me, if I see it approaching and ready to attack.

In each of these cases, we can easily imagine that any evidence that I bring to bear to substantiate my claim will be taken as confirming the hypotheses of the doubters who surround me. The credibility of my testimony will be invariably rejected, as it is expressive of a conclusion that the skeptics will not accept. In this context, anything I say on my behalf is bound to be question-begging. But I am nevertheless justified in regarding my supporting beliefs, and the claims they seek to vindicate, as eminently plausible.

I don't say that our belief in the existence of considerations that oppose the actions of the immoralist is as epistemically secure as the contested beliefs in the examples just given. That is not the point of introducing them. Rather, the examples are designed to show that some question-begging claims are credible and justifiedly held. So even if the various beliefs that condemn the actions of the immoralist beg the question against the instrumentalist, such beliefs might be epistemically justified.

To pursue this path, we would need to distinguish between those question-begging beliefs that are, and those that are not, justifiedly held. I don't intend to embark on such a discussion, because I do not believe that the considerations that support premise (2) are, in fact, question-begging.

I don't believe that only those already convinced of the existence of categorical reasons will find these considerations compelling. What *is* true is that dedicated instrumentalists will find something to resist. I submit that those who have yet to develop a considered view about the existence of categorical reasons will find the considerations that support the second premise

natural and highly plausible. They won't need convincing that there is something to be said against the torturer's actions, and something to be said in favor of easily preventing a child from being needlessly killed. They will then discover that such considerations, when conjoined with an uncontroversial definition of reasons, entail the existence of categorical reasons. The only ones who will deny the appeal of such considerations are those whose theoretical commitments already require them to do so. That such sensibilities are offended is not enough to undermine any success this first argument may enjoy.

III

A second argument on behalf of categorical reasons also relies on our views about the dedicated immoralist, but shifts the focus to matters of responsibility and blameworthiness. Consider those who freely commit themselves to blowing up civilians in crowded areas. Such people are (with perhaps rare exceptions) highly capable of assessing options, gathering information to discover how best to pursue their chosen goals, and taking the needed steps to ensure that their goals are met. They are not insane. They are genuine agents, responsible for their deeds. They are as blameable as agents can be.

We would rescind our condemnation of such people were they literally compelled to do what they did. We would mitigate the blame were we to discover that they had been coerced or manipulated into doing what they had done. But on the assumption that the killers have autonomously elected to proceed in their undertaking, then they are, at the very least, rightly subject to blame.

One is blameworthy for an action only if there is some reason to refrain from committing

it. Because the killers are blameworthy for their deeds, there is a reason that opposes their actions. Since this reason does not depend on the ends that the killers happen to have, the reason is a categorical one. That they have violated or ignored it is the basis of their blameworthiness.

We have here the makings of a second argument for categorical reasons:

1. If one is blameworthy for doing something, then there is a reason not to do it.
2. Uncoerced fanatics are blameworthy for their killings.
3. Therefore there is a reason for these fanatics not to perpetrate such killings.
4. Such a reason, by hypothesis, is neither the content of one of the fanatic's commitments, nor instrumental in securing or protecting one of his commitments.
5. Therefore such a reason applies to the fanatic independently of his commitments.
6. Therefore there is at least one categorical reason.

There are only three premises to the argument, and I think that they are each highly plausible.

Premise (4) is a stipulation that comports with the relevant possibilities, and should be granted by all parties to the debate. It is easy to conjure situations in which perfect instrumentally rational deliberation, begun from a fanatic's existing commitments, would generate no consideration that opposed his deadly undertakings.

Premise (1) seems to me a conceptual truth. If no reasons oppose an action, then those who commit the action are not blameworthy for what they have done. Being deserving of blame entails that one has (at the least) ignored a relevant consideration that opposes the action that one has performed. If one has complied with all relevant and applicable reasons, then there is no room for criticism. And if there is nothing criticizable about an agent's actions, then the person is not properly blameable for his behavior. So a person is blameworthy for an action only if there is some reason that stands against it. That is what the first premise says.

Now consider premise (2). It can be supported thus: if *any* agents are blameworthy for their actions, surely those who are bent on evil are among them. This is so whether the

immoralists are doing evil for its own sake, or doing what is in fact evil, all the while characterizing their actions to themselves as ones that are aimed at a good. An uncoerced, rational fanatic is the perfect exemplar of the blameable agent. His consistency is no proof against criticism. His intelligence and cunning, his ability to select appropriate means to his chosen ends, render him more, rather than less, liable to blame. The standard exculpation conditions do not apply here. The dedicated evildoers are not compelled to act as they do, but have chosen their path and have ruthlessly pursued it in the absence of compulsion, coercion, necessity or factual ignorance.

As far as I can tell, there are only three ways to try to falsify the second premise. The first asserts that fanatics are immune from blame, because their existing commitments will prevent them from seeing the reasons that oppose their misdeeds. If they cannot see these reasons, then they are not blameable for ignoring them. The second way to reject premise (2) is to deny the existence of informed, uncoerced fanatics. The third is to deny that anyone is blameworthy for anything.

The first criticism of premise (2) acknowledges that there can be autonomous fanatics. It acknowledges that some agents are blameable for their misdeeds. But it denies that autonomous fanatics are among them. According to this line of argument, such fanatics are blameless, because flawless reasoning on the basis of their existing commitments will not lead such agents to acknowledge the reasons that oppose their misdeeds. They cannot recognize the error of their ways, since (by hypothesis) none of their commitments will be served by refraining from their evil conduct. And if they cannot recognize their error, then they are free of blame for committing it.

Let us concede a point that is in fact debatable, namely, that flawless practical reasoning is a matter entirely of identifying means to one's endorsed ends. In that case, rational fanatics cannot recognize the existence of reasons to refrain from cruelty, since such reasons, by hypothesis, are not instrumental to their ends. But their inability to appreciate the existence or force of such reasons does not immunize them from blame, if they are blameable for having endorsed their ends in the first place.

To see this, imagine a person who has promised another to meet him at a certain place and time, but then, through her culpable negligence, finds that it is impossible to fulfill the promise. This inability does not cancel her liability to blame. So, too, if the fanatic's prior culpable choices are rendering him unable to see the merits of refraining from his actions, then his inability to appreciate these considerations does not immunize him from blame.

The question thus devolves to one about whether the fanatic's initial choices to ally himself with evil ends are choices that he is blameable for. And why wouldn't they be? I am not imagining a person who has been brainwashed or neurologically manipulated, but someone who makes choices that are as uncoerced and as informed as those of anyone with more ordinary moral preferences. As far as I can see, the only reason to suspend blame here is because one is supposing that no one's choices are blameable. Such a view may be true. But so long as we are willing to blame anyone for the choices that he makes, then we should be prepared to blame the fanatic for his. And that means that his subsequent choices and actions, even if they are endorsed by his instrumentally rational deliberations, are ones for which he is blameworthy. That is just what premise (2) states.

A second criticism of premise (2) claims that autonomous fanatics are not blameworthy

for their actions, because there can be no such thing as an autonomous fanatic. This criticism, sometimes heard in Kantian corners, strikes me as highly implausible. If the claim is more than an instance of a wholesale denial of personal autonomy, then there must be some special reason that agents are unable to autonomously elect evil, though they are able to freely attach themselves to the good. But what could this special reason be? The evil, recall, need not be conceptualized as such—the autonomous fanatic may tell himself that what he is doing is good, and be a dedicated evildoer nonetheless. And he may surely pursue what he really cares about while free of coercion, and in possession of relevantly full information. Certainly, absent clear and compelling argument, we are warranted in abiding by the general maxim that anything apparently conceivable is possible. It seems that we can conceive of the autonomous fanatic. Thus absent a very strong argument to the contrary, we are right to suppose that such fanatics can exist.

A last basis for rejecting premise (2) comes from the assertion that no one is blameworthy for anything. This might be true. If so, my second argument is unsound. I can't say anything here to falsify this potential criticism. All I can do is to express the conviction, shared by almost everyone, that at least some people are rightly blameable for their poor choices and actions. The examples used to substantiate this conviction seem to me more compelling than any of the premises employed in arguments to defeat them.

Because my second argument rests in part on this undefended conviction, it is best to conceive of its conclusion conditionally: *if* anyone is blameworthy for any of her choices, then there are categorical reasons. We get to this conclusion by means of a conceptual truth (premise 1), an uncontroversial statement of possibilities (premise 4), and a highly plausible premise (2)

that expresses a deeply commonsensical assessment of evil-doing.

IV

The most powerful kind of philosophical criticism is one that reveals a contradiction in its target. I have not presented such a criticism of practical instrumentalism, because I do not believe that instrumentalism entails a contradiction. Nor must instrumentalists exemplify any kind of practical inconsistency in behavior or commitment. Most defenders of categorical reasons, following Kant, have tried to sustain such charges. Their vindication would be welcome news for friends of categorical reasons. But I am not optimistic about this most direct route to instrumentalism's refutation.

If a view is not internally contradictory, then any successful criticism of it must proceed by adducing nonclusive but highly plausible reasons, cogently put together to make a strong, albeit defeasible, case. That is what I have tried to do here. Of course, what counts as a plausible reason is relative to antecedent beliefs and commitments. If one is already devoted to instrumentalism, then one will find the considerations I have offered less plausible than anyone else. But that does not distinguish the instrumentalist from (say) the skeptic about other minds. Such people can have internally consistent views, and will regard with great suspicion the supporting evidence introduced by their critics. Still, for those not antecedently wedded to this skepticism, the falsifying evidence can be compelling.

I think that the very same thing is true of practical instrumentalism. We cannot prove that there are categorical reasons. But when we vividly contemplate a world without them, one in which there is literally no consideration that stands against the actions of a torturer, and none in

favor of easily rescuing a child from imminent death, most of us will find that instrumentalism has as much appeal as the various sorts of skepticism that we take seriously only in the study.