It has been a long-standing aim of analytic ethics to provide a formal deduction of our moral obligations from non-moral premises. Nagel (1971), Korsgaard (1996), and Parfit (1987) have all contended that certain facts about the nature of reasons, the will, and the self entail that we have moral commitments to our fellow human beings. While this line of argumentation is undoubtedly valuable in its own right, this paper will attempt to pursue an altogether different approach. Like Korsgaard does in her 1998 article, it will begin with the instrumental principle (Korsgaard 1998). That is where the similarities end, though: rather than seek, like Korsgaard does, to deduce the various sorts of non-instrumental normative commitments that the instrumental principle itself requires, I will instead search for the non-normative facts about the self which are conditions of the instrumental principle’s normative validity. This type of transcendental argument, which moves from a normative (Though in my case, non-moral) principle to the non-normative conditions of this principle’s normative validity is best exemplified by Kant’s third antinomy and the question of what notions of freedom are “necessary and sufficient for practical use” of our reason. In these passages, Kant attempts to argue that the validity of moral principles require transcendental freedom, and I will similarly attempt to argue that the normative validity of the instrumental principle requires certain facts about the self to be true.

By “normative validity”, I mean simply that a principle always gives us reasons which we ought to consider in acting. For the instrumental principle to possess this type of validity, I will argue that agents must A) be rationally capable of regarding themselves as possessing a singular self that reasons, ends, and means all belong to (the singularity condition), and B) that they must be rationally capable of distinguishing themselves from oth-
er selves that possess ends (the distinguishability condition). Furthermore, these two conditions must C) actually obtain for the instrumental principle to be valid (The reality condition), which means that 1) we have a singular self that reasons, ends, and means all belong to, and 2) that selves are actually distinct from one another.

The structure of the paper will be relatively simple. I’ll begin with an examination of the instrumental principle and its characteristics, then give two transcendental arguments. I’ll then discuss the consequences of the arguments I’ve given, and in doing so make a case that the prior two arguments entail that all three conditions hold. The paper will conclude with a consideration of the latent incompleteness of its findings.

The Instrumental Principle

Formulating a rough-and-ready definition of the instrumental principle is surprisingly elusive. One could attempt to define the principle neatly by appealing to a philosophical giant like Kant, but unfortunately there is disagreement over what Kant himself meant by a “hypothetical imperative”. Consider his most direct formulation: “Whoever wills the end, also wills (insofar as reason has decisive influence on his actions) the means that are indispensably necessary to it that are in his control” (Kant 2018: 34). There are commentators who treat this as simply stating an analytic truth about the nature of the will (Finlay 2009), and those who take it to be a normative requirement on action which we can fail to follow (Hill 1970:430). And even amongst those who agree with the latter, there is disagreement over whether the hypothetical imperative has narrow or wide scope. Does it tell us, for example, that we ought to will the necessary means to our end or give that end up (Hill 1970, Darwall 1983), or merely that if we have some end, we ought to will the necessary means to that end? (Schroeder 2005). Both of these accounts run into problems: the former threatens to make the hypothetical imperative into a categorical one (see Schroeder 2005 for this criticism), and the latter runs into counter-examples. Why, after all, must we will the necessary means to our ends when we find out that it’s destructive to our entire system of ends, or immoral?
Therefore, Kant cannot provide us with any perfect formulation of the instrumental principle. Indeed, given that the whole Kantian notion of the “will” is chock-full of metaphysical baggage, it might be better to avoid his formulations and instead speak in the much less controversial language of “reasons”. A reason is, roughly, a consideration in favor of acting in a certain manner. A decisive reason is one that outweighs all the other reasons. There are many formulations of the instrumental principle along these lines (Schroeder 2014, Jollimore 2005, Bratman 2009). But there’s also dispute. Does the instrumental principle counsel that we have reason to do the means that are sufficient for some end, or merely those that are necessary? Should we phrase it as being about ends we have decisive reasons to carry out, or merely those that we have some reason to carry out? Indeed, should we speak of ends in the first clause of the instrumental principle, or merely reasons? Consider two formulations:

Instrumental principle 1: When we have an end, we have a reason to carry out the means to that end.

Instrumental principle 2: When we have a reason to do E, we have some reason to do what will be a means to E.

The latter principle eliminates the notion of an end in favor of having two “reasons”, one in the beginning clause, and one in the ending clause. We can call the former reason the “founding reason”, and the latter a “derived reason”. I’ll call the reason in the ending clause of 1) an instrumental reason. There are differences between these two principles which can be illustrated by some examples. Consider, for example, the case of Jack. Jack is a greedy oil mogul whose sole love is for his eldest son, to whom he decides to leave all of his inheritance. Many would say that he has a reason to give some of his money to the global poor, even if this reason is not related to any of his present desires, beliefs, or intentions. On this account, since he has a founding reason to donate to the global poor, he has a derived reason to find some charity to give to. Therefore the instrumental principle applies to him in formulation 2). In formulation 1), though, Jack doesn’t set any end to donate to charity, and he therefore doesn’t have any instrumental reason. This seems to be right. After all, if Jack is display-
ing some type of irrationality here, it would not be in failing to follow the instrumental principle, but rather in failing to live up to his moral demands. We would only say he was failing to follow the instrumental principle if he had already decided he had reasons to give to the global poor (or set an end to give to them), and never gave to charity. The proponent of the second principle might object that this depends on an externalist account of reasons, and that so long as we suppose that the founding principle was internal, we would be fine. I will use Kieran Setiya’s paraphrase of Williams’ classic position to define internalism: "P is a reason for A to x if and only if there is a sound deliberative route from A’s beliefs, taken together with his subjective motivational set and the belief that P, to the desire to x" (Kieran 2004). Now consider Jackie, who has a great zest for learning. Her happiness and fulfillment would best be served by her becoming a professor. However, because she has not reflected on the matter at any length, she believes she will be better served by going into law. She has thus never set as her end the decision to become a professor. In this case, it is clear that Jackie has no instrumental reason to, say, take the GRE’s in order to become a professor, even if it would be what best serves her long-term interests. We would not judge her instrumentally irrational in failing to take the GRE’s. It seems, however, that principle 2) implies that we ought to. So that’s another count against 2). There is, however, a trivial sense in which 2) is true. If we have a reason, whether external or internal, to do some action, then we do indeed have a reason to carry out the means to that action. But this isn’t a version of the instrumental principle, since it doesn’t really make sense to say an individual is instrumentally incoherent in failing to follow it. So we ought to adopt 1). Of course, there are other disputes that might be pertinent. One might wonder whether it’s better to add the prefix “necessary” or “sufficient” to means. And of course, the definition writ large could be made more complex to rule out various counter-examples. But I’m going with a simple definition here, because I want to show that even with the most primitive notion of the instrumental principle, fairly significant conclusions can be deduced.

It’s also worth noting two key elements of this new definition of the instrumental principle: the transmission relation, and
the necessary structure of that relation. The former can be made evident by examining the internal structure of the instrumental principle. We only have an instrumental reason to perform means because that means is related in a proper way *qua* means to a given end. There might be some other reason to perform some means (say, if that means was the right thing to do in some specific situation), but that would not be an *instrumental* reason. It is only because of this relation to ends - as things we value, desire, or have reason to want - that we have reason to engage in actions we would qualify as ‘means’. That is why we speak of a “transmission”: by virtue of the relationship between ends and means, the end transmits to the means a reason to perform it. This should *always* happen. If it doesn’t, the problem will be with our phrasing of the instrumental principle, not with the instrumental principle itself.

This is why there is a necessary structure to the transmission relation. Although it is contingent matter that we have adopted a certain end, it is a matter of necessity that once the ends and means are determined, the ends transmit to the means a reason to perform them. If this was not a necessity, then it would seem to be a contingent, empirical circumstance that once the ends and means are fixed, the ends transmit to an agent a reason to perform a means, which would be simply bizarre. This would mean that one could give counter-examples to some given version of the instrumental principle, and a suitable rebuttal would be to merely point out that the instrumental principle can of course fail to transmit reasons from ends to means, since it isn’t necessary. But that’s not what the counter-examples show: they demonstrate, rather, that some phrasing of the instrumental principle is insufficient since ends don’t necessarily transmit reasons to means.

This allows us to rephrase the instrumental principle in the following manner: “When some x has an end, that same x necessarily has a reason to carry out the means to that end”. In the following two arguments, I will use this formulation of the instrumental principle. I will also use the variable ‘X’ instead of our generic concept of a person, because less metaphysical baggage comes with it.
The Ownership Argument

Note an interesting formal fact about the instrumental principle: it is only valid if a given X possesses various ends. X having a reason to do end \( a \) gives Z no instrumental reason to do \( b \) if it will serve as a means to end \( a \), unless Z already has end \( a \). Of course, Z might have a moral reason to set end \( a \) as their own, but this would not give them the instrumental reason to will the means \( b \) to end \( a \) unless they have already adopted end \( a \). More concretely, if Jack wants to go biking, Jill has no instrumental reason to buy Jack a bike unless she is already in some way concerned with Jack’s pursuits, or decided to act on a moral consideration. In other words, Jill has no instrumental reason to buy Jack a bike unless she already has set an end that will be fulfilled by buying Jack a bike. Another example, with less of an explicitly moral valence, might be more illustrative. Consider Jackie. Although she would be best served by becoming a professor, she has decided she wants to enter law. Without having set the end to become a professor, she has no instrumental reason to take the GRE’s, even though it is what best serves her interests. This demonstrates that for an X to have any particular instrumental reason, that same agent must have some end.

Another way to speak of this is in terms of the transmission relation. Instrumental reasons get their normative force from their relation to ends. It is not enough, however, that these ends be just anybody’s. These ends have to be one’s own in order for them to grant instrumental reasons. As we have seen with both Jackie and Jill, their moral and prudential reasons to set various ends do not grant them analogous instrumental reasons to actually accomplish those ends unless they have actually followed the counsel of these principles and set various ends as their own. One can, of course, consider everyone else’s end as one’s own (the Kantian does something like this in the Kingdom of Ends), but this serves to exemplify, not nullify, the requirement that we must have ends as our own in order to deduce instrumental reasons for them.

For the transference relation to go through, then, an X must be capable of conceiving of an end as their own. And to conceive of an end as one’s own is to conceive of it as belonging to you, rather than some other X. What this “you” or “self” is we
cannot say at this point, but it seems fair to claim that this strange X possesses ends that can be distinguished from other X’s that have ends. This is not to say that the ends must be distinct (indeed, in the case of the Kingdom of Ends, they might be precisely the same), but the thing that possesses these ends must be conceived of as distinct if the instrumental principle is to go through. If we were not capable of distinguishing the thing that possesses these ends from other things that possess ends, we would not know when a given end is our own or another’s, and thus the instrumental principle would not be valid since we could not confidently say when a given end is our own. And if we could not confidently say when an end is our own or another’s, the instrumental principle would not necessarily transmit reasons from ends to means. There’s also a conceptual argument to be made: it would be entirely insensible to say of some end that it was your “own” if you were not readily distinguishable from other selves. In that way, the very concept of “ownness” presupposes the distinguishability condition, because without the capacity to distinguish ourselves from other selves we wouldn’t be able of knowing if an end was “our own” or another’s.

The Unity Argument
The instrumental principle explicitly involves three essential elements: means, ends, and reasons. Let us say a means is something some X must do in order to achieve an end. Let us denote this X as an ‘actor’. The foregoing argument demonstrates that we must have ends that are our own if the instrumental principle is to be normatively valid. Let’s call the X which possesses, or conceives of ends as its own the ‘agent’. Now we can formulate the instrumental principle as “Whenever some agent possesses an end y there is reason for an actor to do whatever is a means z to that end”. There is a striking problem with this formulation: if the actor and the agent are conceived of as having different referents, there is no reason to suppose that the transference relation necessarily holds between end y and means x. If the actor was not identical to the agent, then the actor might not possess the ends that the agent possesses, which means that it would not necessarily have any instrumental reason for doing some means since it would not be a means to its own end. Thus the actor that is
viewed as having a reason to do some means must be regarded as numerically identical to the agent that has an end.

We can replay this argument by considering the X that has reasons. Some might doubt that any X ‘has’ reasons. If we do assume internalism (about reasons, not principles), reasons must be capable of having some degree of motivational force. They must, in Korsgaard’s words (Korsgaard 1996: 85), be capable of “getting a grip” on some X. A reason must be a reason for some specific X if it is to have the requisite motivational force. For some philosophers ((Parfit 1987) and (Scanlon 2000)), though, reasons are simply things that are out there, waiting to be discovered. But even in this case, it is clear that these reasons must be capable of being relevant for various X’s. Let us call this x that is capable of having reasons be relevant for it an ‘author’. Now we can rephrase the instrumental principle “Whenever some agent has an end y, there is a reason x for an author to do whatever is an actor’s means z to that end”.

It ought to be abundantly clear that if we do not conceive of the agent, the author, and the actor as numerically identical, the instrumental principle will be burst asunder. The entity that the reasons are relevant for must be regarded as the same entity that possesses ends and that is capable of having means to certain ends. If the author that has reasons was different than the agent that has ends and the actor that has means, then it would not be necessary that the agent and actor have the same reasons as the author, since they would be different entities. But this clearly is necessary for the instrumental principle to hold: if the instrumental principle did not give the actor or agent reasons, then the instrumental principle would not be normative for them at all, since it would not provide them with actual reasons that tell them that they ought to act in a particular manner.

We can also flip this around and consider matters from the perspective of the author that has reasons, but is non-identical to the agent that has ends and the actor that has means. In this case, they would have certain reasons, but we couldn’t call these reasons properly instrumental since they would not necessarily serve to fulfill its own ends; they might have altogether different ends given their non-identity with the agent. But this necessity clearly is a part of the instrumental principle, as we’ve argued earlier. Therefore the entity, whatever it is, that responds to reasons must be conceived as the same entity that has ends. Otherwise, this X
that responds to reasons would not necessarily have a reason to fulfill the ends of the agent, since it does not necessarily possess these ends as its own, which is a requirement for the validity of the instrumental principle. Thus we can say that for the transference relation to hold, we must regard the same X as responding to reasons, having ends, and possessing means to those ends. Let us denote this a ‘self’.

What the Two Arguments Show

There are two possible objections to the arguments I’ve given thus far. One line of response is that what I’ve shown is merely how we must regard ourselves if we are bound by the instrumental principle, since transcendental arguments serve only to illustrate what we must believe to be true and not what is actually the case (Stroud 1968: 255). This comports well with an anti-realist theory of principle validity, in which principles are normatively valid if agents hold certain kinds of attitudes and beliefs. In this case, what I’ve shown is just what configuration of beliefs and attitudes are necessary for the instrumental principle’s validity - but I haven’t shown that those beliefs must actually be true.

The other response pushes in the other direction. Isn’t it the case that the instrumental principle is binding on us regardless of whether we consider ourselves to be unified and distinguishable selves? On a realist account of practical principles, that certainly might be right. Let’s say a realist account of practical principles is one that holds that the desires, beliefs, and attitudes of any given agent are totally irrelevant to the validity of a given practical principle. Rather, what makes a certain normative principle valid is certain non-normative, agent-neutral facts that serve to “ground” a principle. Let’s say that a proposition “grounds” a principle if the principle will be normatively invalid if the proposition is false. In this case, my arguments can be interpreted as offering a proof that the instrumental principle is grounded by various propositions. After all, if we did in fact not have unified and distinguishable selves, it would be very difficult to make sense of the instrumental principle’s validity on a realist account. For the transmission relation to be necessary, there needs to be an identity between the X that has reasons, ends, and means. Similarly, for agents to have instrumental reasons to act on their ends,
these ends need to be their own and not of others, which means selves need to be distinguished from each other. If either of these conditions did not actually obtain, the instrumental principle wouldn’t be valid, since the transmission relation would break-down. The realist account thus allows us to effectively fast-track our way to the reality condition.

What about the singularity and distinguishability conditions? These hold merely that agents must be capable of distinguishing themselves and holding themselves to be singular selves. It is thus unrelated to the actual beliefs and desires of any current agent, and hence is perfectly acceptable on a realist account. And I think, too, that there are compelling reasons for a realist to want to accept these conditions. Let’s say that a principle “presupposes” a belief if the agent must hold that belief if they are to rationally accept the principle as normatively valid (regardless of whether it actually is). Hopefully, I’ve shown that the distinguishability and singularity conditions are presupposed in this way. If we are to rationally hold that the instrumental principle has normative validity, we must in turn view ourselves as having unified and distinguishable selves. Now consider the principle “ought implies can” (hereafter OIC). Following Johnny Anomaly’s article on the philosophical forum, OIC claims that no act “can be morally required if it is beyond human capacities to perform” (Anomaly 2008: 480) Reformulated to apply to principles, OIC claims that no principle can be normatively valid if it is impossible for human beings to rationally regard the principle as being valid. In other words, if we cannot regard it as being valid, it cannot be valid, since it could never be normative for us. For a principle to be normatively valid, it must be capable of rationally guiding human action (Anomaly 2008:480, from Griffin 1992) - but if we cannot rationally accept the grounds on which that principle rests, then the principle will not be capable of rationally guiding our action.

If the realist rejected the distinguishability and singularity conditions, they would in effect be rejecting ought implies can. If agents were not capable of viewing themselves as singular and distinguishable selves, they would not be capable of rationally regarding the instrumental principle as valid. This would entail the bizarre position that the instrumental principle is normatively
valid yet simultaneously agents are incapable of rationally regarding such a principle to be valid. While I can’t say this is surely incorrect, I think our intuitions (as well as the principle of ought-implies-can) speak strongly against it.

How about the anti-realist position? On this account, I’ve shown merely that we need to accept certain facts to believe that the instrumental principle is normatively valid, but not that these facts must actually be true. The best argument I can give against this position is by example. Let’s suppose that the prudential principle ‘presupposes’ the existence of a continuously existing self. Now assume that such a self can be demonstratively shown to not exist, and that Johnathan labors under the delusion that it does. Is the principle of prudence normatively valid for Johnathan? It makes most sense to say that he might regard this principle as normatively valid, but that it lacks real normative force because it is based on false assumptions. Another example might make this more lucid. Suppose that moral principles require the existence of other agents who feel pain, but there are in fact, no agents like that which exist. Johnathan lives in a solipsist world, where his peers are all elaborate automatons. But he mistakenly believes them to be real persons. Is the moral principle truly normatively valid in this case? Johnathan might believe it to be normatively valid, but surely it is not a principle that he ought to act on, all things considered. And let’s consider, lastly, our own case. There is no self that has ends, means, and responds to reasons. At any given moment, we are inhabited by multiple selves, one of whom has an end, and another whom has a reason. If an agent erroneously labors under the belief that this is not the case and employs the instrumental principle, they are doing so by a sort of mistake. In this case, the instrumental principle is clearly not normatively valid, though an agent might mistakenly regard it to be so. Therefore, for the instrumental principle to have normative validity, the beliefs which are presupposed by it must be true. This establishes the reality condition and concludes our proof of the transcendental presuppositions of the instrumental principle’s validity. For we have shown on both realist and anti-realist accounts of principle validity that the instrumental principle is valid.
The Nature of the Self

This paper is incomplete. It specifies several conditions that must be fulfilled for the instrumental principle to be valid, but does not explain how they are to be fulfilled. There is a massive literature on self-reference, personal identity and self-consciousness I have neglected to engage with here (see Castaneda 1966, Shoemaker 1968, Perry 1979, and Kaplan 1989, bibliography of Gertler 2015, Olson 2015, and Smith 2017) which might explain how the various conditions might be true. That is a project worth pursuing.

There’s another sense in which the argument of this paper is incomplete. So far, I’ve examined only the presuppositions of instrumental reason. I began with the thinnest principle of rational action because its validity commands the broadest philosophical assent. If Korsgaard is right, however, that the instrumental principle cannot be the sole requirement on practical reason, then our analysis here is drastically incomplete (Korsgaard 2008). An examination of the transcendental conditions of the instrumental principle must give way to an investigation of the conditions for prudential and moral principles. Of course, this project is far beyond the scope of this paper. But it is one that I hope I have provided a preparatory sketch for in the present work.
Notes

2. Gareth Evans’ notion of fundamental identification is close to this (See Roos 2004).
3. This serves to demonstrate that prudential is a different species of practical rationality than instrumental reason. Nagel argues against the opposing view in The Possibility of Altruism (2008).
4. See Kiesewetter’s article (2015) on the transmission relation for a defense of the very specific form of the transmission principle that we rejected earlier.
6. On the Kantian formulation, it involves means, ends, and a will. On other formulations, it involves ends, means, and an “ought” statement. This is how Mathew Bedkey (2009) formulates it. Ultimately, this latter formulation is not very different from ours, since an “ought-statement” can be generally translated into a “reason-giving” one.
7. For a response to Korsgaard’s arguments against externalism, see Shaver (2006).
8. It is worth distinguishing between the ontological status of reasons and their normative status. That is, reasons might very well be the sort of things that can have normative command over us only if they can motivate us, but also exist, in some sense, independently of agents. Conversely, reasons can have unrestricted normative command over us as agents but also be dependent on the existence of agents for their existence. I think this latter view comes closest to the Kantian one, though Korsgaard would, of course, dispute this. It’s also clear that it’s at least potentially possible to argue from normative internalism or externalism to their ontological correlate. This relation of logical entailment doesn’t mean that the positions are identical, though.
9. To see a response to Stroud’s article against transcendental arguments, consult (Stern 2000).
References


