A New Distinction in Meta-Ethics

David DeMatteo

Introduction

The purpose of this paper is to make a new distinction in meta-ethics. Specifically, I will distinguish between externalism and internalism about normative principle validity (hereafter EINP). Given that the whole internalism/externalism schema has been applied to matters as diverse as mental content, epistemological justification, moral judgment, and reasons[1], it might seem as if there's no need to make further use of what is already becoming an overworn trope. But in this essay, I'll argue that my new employment of the distinction is not only conceptually original, in the sense that it isn't reducible to any other uses of the distinction, but also useful for clarifying our thought about normative principles. By normative principle, I mean some proposition that states that we have reasons to act in some particular manner. For example, the hypothetical imperative "You have reason to do whatever fulfills your ends" is a normative principle, since it provides us with reasons given that we have certain ends. "You should help others in need" is also a normative principle, even if it doesn't ex-

plicitly include the language of reasons, since it tells us that in certain situations we have reason to engage in certain types of actions. Normative validity, meanwhile, refers to whether a principle serves as a normative guide on our conduct by generating reasons which we ought to consider in action (Korsgaard 2008, 31, Korsgaard 2014, 80). This is different from a more robustly realist conception of validity which holds that some principle can be valid regardless of whether it generates reasons which we ought to consider in action. Principles, in this picture, are always normatively valid for some agent or set of agents, and what we will ultimately be searching for in this paper is what the conditions are for a given normative principle to be valid for some agent. A reason, meanwhile, is simply a consideration which counts in favor of acting in a certain manner (Nagel 2008). Note that these reasons don't have to be decisive: they might be outweighed by other reasons.

By making this new distinction, I mean to illuminate a certain conceptual space in which one can stake out various claims. When philosophers distinguish, say, between mental content externalism and internalism, they mean to point to a spectrum of various positions that can be taken about how the contents of our mental acts are constituted. Similarly, by employing EINP, I will attempt to show how a range of philosophical theories can be mapped along two opposing poles. This will only

be a useful device, of course, if the distinction I'm making is conceptually novel (I.e, picks out a real distinction) and is hence irreducible to a variety of other schemas which aim to accomplish a similar task. Therefore, the first portion of this paper will be dedicated to providing an exposition of EINP and show that it isn't reducible to reasons internalism/externalism, moral judgment internalism/externalism, and realism and anti-realism about ethical propositions. With the basic conceptual originality of the distinction in place, I'll then advocate one particular position along the principle internalism/externalism spectrum, which I call the convergence position. Finally, I'll conclude by making some closing remarks on the philosophical worth of the distinction by both situating it within a broadly Kantian problematic and showing how it can help resolve certain disputes in metaethics.

The Distinction

Are normative principles valid for some agent because of beliefs, dispositions, and attitudes that agents have, or are they valid because of certain facts about the world? This is the basic issue around which EINP revolves. To the close reader, it might seem like this question creates a dichotomy where none exists: aren't facts about agents also facts about the world? We should distinguish agent-neutral facts and agent-relative facts. Agent-neutral facts are those which pertain

to all agents, like, say, the fact that they are agents and have desires. Agent-relative facts are those which are only true of certain agents, like the fact that I have a desire to someday enter the philosophy profession. So now we can clarify our initial question: are normative principles valid for an agent because of particular beliefs, dispositions, and attitudes that individual agents have, or are they valid because of certain facts about the world and agent-neutral facts about agents? Some examples are probably in order. Agent-neutral facts are those like the fact that we have desires and are capable of deliberating about what we ought to do. Kantian moral theory holds that general facts about practical reason and the nature of agency can be used to deduce principles which are binding on agents (or normatively valid). Notably, for Kantian moral theory, it is an agent-neutral fact that to act is always to act according to some maxim. This fact about the nature of agency is used by the Kantian to argue for the necessity of universal law (see Korsgaard 1996). Agent-relative facts are those like the fact that I have a desire to help others. On certain self-interest theories of a Humean variety, a principle claiming I ought to help others would not be normatively valid for me unless I already possessed such a desire.

By saying that principles are valid for some agent because of either particular beliefs and dispositions or

agent-neutral facts, I don't mean to imply that there is some causation going on here. "Because" merely indicates that there's some relationship of dependence: in an externalist picture, a principle might not be valid if certain facts about the world are not true, but that doesn't entail that the truth of those facts cause the principle to be valid. They are rather conditions for its normative validity, and our inquiry here is really into what the conditions are for any given normative principle's validity. Are these conditions basically bound up with the particular attitudes of agents, or are they dependent on more general facts about the world?

Internalist positions hold that the normative validity of principles is dependent in various respects on the propositional attitudes of particular agents. An extreme internalism about principles will thus hold that a principle is valid for an agent by the mere fact that an agent regards it as being valid. A more moderate internalist will assert that there is a complex web of beliefs which are necessary for a principle to be valid, but it is still ultimately a matter of an individual's attitudes and beliefs. In this case, the agent might need to have certain beliefs in not only the validity of the principle, but also other propositions which are rationally entailed by the principle. Or they might have to simply rationally believe in the principle's validity, and not hold that it is valid merely because of some personal idiosyncrasy.

That sets up constraints on the internal configuration of beliefs which can vouchsafe a principle's validity, but the validity of those principles is still wholly dependent on agent-relative facts.

By contrast to these internalist positions, an extreme externalism claims that a normative principle's validity does not depend on the beliefs and attitudes any given agent has, and can be valid for a particular agent even if it is impossible for that agent to rationally consider it as valid. The only criterion for its validity is that the facts about the world which are required for the principle's validity be true. There is also a weak externalist position which is possible to stake out in this conceptual space, which holds merely that the current configurations of beliefs and desires that any agent has cannot on their own be a sufficient condition of the principle being valid for that same agent. In other words, some agent-neutral fact (or fact about the world) must be true for the principle to be valid. Corresponding to this form of weak externalism is a weak internalism, which asserts that subjects must be at least capable of rationally regarding the given practical principle as being normatively valid, which would also entail, of course, holding any beliefs which are entailed by the principle. Note that weak internalism and externalism are not exclusive, but rather eminently compatible with one another. We'll call the fusion of these two positions the convergence po-

sition. Later, I'll provide a limited argument in defense of it. For now, though, we need to defend EINP itself.

A Defense of the Distinction

The general distinction between internalism and externalism has occasioned fierce debates in a wide swathe of philosophical sub-fields. As far as I can tell, though, nobody has yet applied it to practical principles themselves, and so it's worth clarifying why the distinction made here is genuinely different from 1) judgement internalism/externalism, 2) reasons internalism/externalism, and 3) realism and anti-realism about normative claims. Judgement internalism/externalism concerns whether moral judgments necessarily provide us with motivations for action.[2] A judgement internalist holds that whenever I make some moral judgement, it will provide me with a motivation for action. If I decide that it is wrong to consume animals, I will therefore have a motivation to stop consuming animals. Judgement externalists hold the opposite position: it's possible for me to judge that meat-eating is immoral without having any motivation whatsoever to cease eating meat. I'll argue that EINP has a different domain from judgment internalism/externalism by demonstrating that the two distinctions don't neatly map onto one another. A philosopher could easily hold, for example, both moderate internalism about moral principles and extreme externalism about moral judgements. If this is the case, the

validity of our moral principles will ultimately be tethered to the particular desires of agents, but the moral judgements we make using those principles won't ever provide us with motivation on their own. This is quite possible, precisely because what makes a normative principle valid is quite different from how that principle serves to both motivate moral judgements and how those judgements in turn interact with our particular attitudes. We can also flip this around, and consider a philosopher who holds weak externalism about principles and moral judgment internalism (A given moral judgment necessarily provides a motivation for the agent to act on it). In fact, this is arguably the position of Thomas Nagel in *The Possibility of Altruism* (2008), who combines a robust moral realism rooted in the external objectivity of moral principles based on agent-neutral facts with an internalism about moral judgments. This is a tempting position for any moral realist who wants to hold that morality's principles are unconditionally binding on individuals due to non-agent relative facts, and that individual moral judgments provide agents with the motivations to follow them.[3]

The debate about reasons' internalism and externalism has a closer connection to practical principles, but, as we'll see, the two distinctions still legislate over different domains.[4] The crux of the difference between internalists and externalists about reasons concerns the

relation between reasons and motivational facts.[5] Internalists hold that for something to be a reason for an agent, it must be tied in some way to their motives and desires (Arkonovich 2013, 210). Externalists hold the opposite: something can be a reason for an agent even it doesn't bear any relation to their propositional attitudes. A reasons internalist will argue that a maddened serial killer has no reason to cease his murders if there is not any connection between this reason and their present motives and desires. An externalist will hold the opposite: there's reason for them to cease their murders even if there is no deliberative route which might connect their current desires and motives to the moral reason to cease killing. Externalists don't have to hold that all reasons are external – they're merely committed to the proposition that some reasons are not internal.

As we'll see, it's possible to hold an internalist or externalist position in one domain without doing so in another. The instrumental principle is illustrative here. Roughly, the instrumental principle states that if we have an end, we have some reason to pursue the means to that end. Now note that I might be some form of an externalist about the instrumental principle (there must be some non-agent relative facts that the instrumental principle depends on to be valid), while also cleaving to an internalist account of instrumental reasons (instrumental reasons must be capable of motivating an agent

for them to be reasons). This account of the instrumental principle is actually quite appealing, because it explains why it is binding on all agents while simultaneously providing reasons that always must be capable of motivating an agent! Any account of the instrumental principle that made its reasons "external" and thus incapable of always potentially motivating agents would be quite bizarre – but we also want to understand the principle itself as having a sort of validity which is due to facts that don't just pertain to particular agents (Bedke 2009, Jollimore 2005). We can flip this around in the case of moral principles, and imagine a case where I am a weak internalist about some moral principle, but also an externalist about certain reasons. In this case, I will believe that this moral principle's validity must be capable of being recognized by an agent for it to be valid for them, but also think that certain reasons might bind those agents regardless of the desires and attitudes which they might have. That is eminently sensible, and in fact might be the best account of moral principles there is: We might want to say that someone has reasons to respond to some principle even if they had no way of being motivated by them, but we also might be doubtful that a principle could be valid for them if they had no way of recognizing it as valid.

Think, for example, of the moral principle: "Whenever it is not unnecessarily burdensome for you, help others".

Now suppose that for some agent, they have no desire or set of motivations to actually aid fellow human beings. In this case, we might say that they have an external reason to follow the normative principle in question, but that the reasons generated by the principle itself are internal because they are only reasons for the agent if they are capable of recognizing the principle's validity. We would thus be externalists because about reasons because we hold that there are some external reasons but also be internalists about principle validity.

Once we approach extreme internalism and externalism about principles, matters do get somewhat muddier. It would be very odd to be, say, an extreme internalist about moral principles and also be an externalist about the reasons that those moral principles give us. If all that is necessary for a principle to be valid is that an agent regard it as valid, then it would be difficult to understand how the reasons that such a principle provides could be externally binding on individuals regardless of their attitudes. It is also difficult to combine an extreme externalism about moral principles (the validity of a moral principle does not depend on any attitude, motivation, or belief an agent has, and would be true even if they were not capable of regarding it as valid) with reasons internalism. Presumably, if the validity of a moral principle didn't depend on any agent-relative facts, then the reasons it provides wouldn't either. Therefore, there are

some relations of entailment between these two distinctions. But this doesn't mean that the two distinctions aren't, in fact, distinct - especially because the entailment relations are incomplete.

Lastly, it's worth saying something about why EINP isn't simply reducible to realism and anti-realism about normative facts. There are three responses we can give, based on three renderings of what it means to be a "moral realist". First, suppose we treat moral realism and anti-realism as mapping onto principle validity itself. In this picture, normative realists are those who hold that normative principles can give reasons, and anti-realists are those who argue the opposite. In this case, we can distinguish between the two positions by noting that EINP assumes realism about normative validity, and then inquires into the conditions for a given principle to have that validity. If we treat moral realism and anti-realism as defined by their positions on the mind-dependence of moral facts, then we can show that it is possible to be both a moral realist and principle validity internalist. One might hold, for example, that moral principles exist in some sense independently of us, but only have normative validity if various conditions hold true which are agent-relative. Christine Korsgaard has pointed out that even if moral facts exist, they need some way of "getting a grip" on us (Korsgaard 2014). In other words, even if there exist moral principles irrespective of us,

that would not imply that those principles are necessarily normatively valid for us. And it might just be that for those moral principles to be normatively valid, they need certain agent-relative facts to also be true. This opens up the path to a fairly deep rift between moral realism and principle externalism. Conversely, one might take the position that it is possible for moral principles to be completely dependent on various facts about agents and have no existence apart from agents, but have those facts be solely agent-neutral ones. In this way, moral principles would be "mind-dependent" in the sense that they are constituted by various truths about agents, but their normativity would still be "externalist" insofar as they are not dependent on agent-relative facts. This would effectively combine moral anti-realism (at least if anti-realism is understood asserting that moral propositions are mind-dependent) and principle validity externalism. Lastly, we can think of moral realism as involving two claims: the capacity of moral judgments to be true or false, and the truth of most ordinary moral judgments (See Sayre-McCord 1986, from Joyce 2015). Just as internalism and externalism about principle validity presupposed the validity of principles, so it presupposes the capacity for moral judgments to be true. Consider if some normative principle was untrue: in that case, it would not provide us with reasons for acting. It would thus be of no use inquiring into its validity, since it would have no validity. If we are going

to inquire into the conditions for the normative validity of some principle, we thus must assume realism. This parallels the earlier case, where the normative validity of a principle was itself a condition for EINP rather than a distinction that could be equated with it.

A Defense of the Convergence Position

Earlier, I detailed what I called the convergence position. Recall that this position holds both A) that agents must be capable of rationally regarding some principle as being valid, which means holding various beliefs that a principle entails and B) that there must be some facts which are not agent-relative if the principle in question is to be true. Note that this is merely a thesis about the necessary conditions of principle validity – it does not state what facts are sufficient for a principle's normative validity, which is a much trickier task (and one that I will refrain from in this paper).

Let's first consider A). This basically amounts to the thesis that if a particular agent has no way of rationally regarding a principle as being valid, it can't be normatively valid for them. The negation of this position is that a principle could be normatively valid for an agent even if they can't rationally regard it as being valid. This is, in effect, an extreme form of externalism, which can be dismissed solely on the basis of the principle of ought implies can (hereafter designated OIC). Following John-

ny Anomaly's article, OIC claims that no act "can be morally required if it is beyond human capacities to perform." (Anomaly 2008) 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 valid. In other words, if we cannot regard it as 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 - 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.

Now let's consider B). This is the claim that if some given principle is to be true, there must be some non-agent-relative facts that are true. Its converse is that there do not need to be any agent-relative facts that are true for a principle to be normatively valid. There are a few ways to stake out this position: one might, firstly, be an extreme internalist who holds that a principle could be made true solely by an agent holding it to be true. In this case, the agent in question doesn't need to have any other beliefs or attitudes except one endorsing the truth of the principle. This would entail the rather bizarre position, though, that all sorts of principles are normatively valid, such as "whenever I can, I will harm myself and others" - even if the agent also believes that pain is objectively wrong. To avoid conclu-

sions like this, one might endorse a moderate internalism, which requires an agent to have a sort of rational consistency in their beliefs that requires believing in any propositions that follow from the validity of a principle. Let us call these beliefs "presuppositions". Take, for example, the prudential principle, which states that one has some reason to do what is in one's future self-interest. Suppose that this 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 presuppose the existence of other agents who feel pain, but there are in fact, no agents like that which exist, and they in fact cannot 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. The principle is only mistakenly believed to be normatively valid.

The above cases demonstrate that the various presuppositions of a principle must also be true if a principle is to be normatively valid. If those presuppositions assert the truth of agent-neutral facts, as they likely will, then the form of weak externalism we've already discussed will be true, since the given principle's validity will depend on various agent-neutral facts. This establishes both the clauses of the convergence position, and thus demonstrates that it is true.

The Philosophical Merit of the Distinction

How tethered must morality be to human life and human practices? This distinction basically speaks to that question. Rather than asking about reasons or moral motivations, however, it inquires into the validity of normative principles themselves. Can a normative principle be normative for some human being if they found themselves incapable of recognizing its validity? Are these principles rooted in deep facts about the world, or merely in the particular desires of agents? These are the questions that this distinction grapples with.

It is concerned with what is basically a Kantian problematic: what are the conditions for the validity of various ethical principles? In answering this question, one must make a transcendental argument[6] which moves from the normative validity of a given principle to the conditions for the validity of that same principle. The

distinction itself thus encourages moral philosophers to make use of transcendental argumentation, which in recent years has made something of a come-back in analytic ethics with the work of Korsgaard (*Creating the Kingdom of Ends* 1996). And of course, we can ask whether the various conditions discovered by transcendental argumentation are rooted in 1) the idiosyncrasies of various subjects, 2) the structure of the world subjects live in, or 3) the *a priori* conditions of agency itself. Notably, those who take these last two positions are both put in the externalist camp, but they might have quite different philosophical predilections: whereas many of the former consider themselves realists who believe moral facts exist on the same plane as platonic mathematical entities (See Parfit 2011), those in the latter group often believe that moral principles are valid because of certain facts about the structure of rational agency (Korsgaard 1996).

If this distinction has merit, it will not just be, though, in its responsiveness to perennial philosophical questions. Its value must also lie in its aim to clearly explicate how a variety of meta-ethical stances can be mapped, and to clarify our philosophical discourse in doing so. For example, there has long been a dispute over Korsgaard's claim that normative claims need some way to "get a grip" on us (Korsgaard 1996). Yet this distinction makes clear that this is a different manner from the

whole question of whether normative principles exist independently from us. One might be a realist about the ontological status of normative principles but an internalist about their normative status. That is, normative principles might very well be the sort of thing that require some connection to our propositional attitudes to have validity, but also exist independently of agents. EINP thus opens up the philosophical space to notice that ontological status and normative status are indeed distinct, and thereby enables us to increase the sophistication of our philosophical discourse. This might be invaluable to philosophers who wish to hold that various sorts of normative principles do indeed possess a sort of mind-independence, but who also don't want to forsake their connection on a normative level to motivational states. At its best, then, this distinction might serve as a heuristic tool, enabling philosophers to more thoroughly clarify where they stand. And insofar we conceive of philosophy as itself a practice of conceptual clarification, the art of making such distinctions is not merely an aide to the practice of philosophy, but itself a form of philosophical practice.

Notes

1. For a paper that nicely reviews and criticizes externalist accounts of mental content, see (Farkas 2008). For a review of the literature around epistemic justification, see the bibliography of (Bonjour and Sosa 2008).

2. In this I am following Russ Shafer-Landau's early definition of moral internalism in "A Defense of Motivational Externalism" (2000: 270). Shafer-Landau also offers an overview of the literature around moral internalism and externalism.

3. Of course, there is an anti-realist argument that depends on motivational judgment internalism, outlined on page 121 Shafer-Landau's 2003 book on moral realism. It moves from an acceptance of motivational judgment internalism and motivational humeanism (beliefs do not yield motivational states) to a robust anti-realism. The key fact here is that motivational judgment internalism alone does not produce an anti-realist position. And indeed, several philosophers who are moral realists of various stripes have rejected motivational Humeanism but accepted judgment internalism, like McDowell (1978) and Scanlon (2000).

4. For some defenses of reasons internalism, see Williams *Ethics and the Limitations of Philosophy* (1985), and Cowley "A New Defense of William's Reasons-Internalism" (2005). For an overview of internalist positions, see Arkonovitch's "Varieties of Reasons/Motives Internalism" (2013). For some broad-sides against internalism, see Setiya's "Against Internalism" (2004), Parfit's "Reasons and Motivation" (1997) , and Brewer's "The Real Problem with Internalism about Reasons" (2002).

5. I take this definition from (Finlay and Schroeder 2017).

6. I.e, an argument that moves from the validity of some given principle or practice (experience, knowledge, etc.) to the conditions for the possibility of its validity.

References

Anomaly, Johnny. 2008. "Internal Reasons and the Ought-Implies-Can Principle." *The Philosophical Forum* 469-483.

Arkonovich, Steven. 2013. "Varieties of Reasons/Motives Internalism." *Philosophy Compass* 210-219.

Bedke, Mathew. 2009. "The Iffiest Oughts: A Guide of Reasons Account of End-Given Conditionals." *Ethics* 672-698.

Brewer, Talbor. 2002. "The Realm Problem with Internalism about Reasons." *Canadian Journal of Philosophy* 443-473.

Copp, David. 2007. *Morality in a Natural World: Selected Essays in Metaethics*. Cambridge University Press.

Cowley, Christopher. 2005. "A New Defence of Williams' Reasons-Internalism." *Philosophical Investigations* 346-368.

Farkas, Katalin. 2003. "What Is Externalism?" *Philosophical Studies* 187-208.

Finlay, Stephen, and Mark Schroeder. 2017. *Reasons for Action: Internal vs. External*. August 18. Accessed November 22, 2018. plato.stanford.edu/entries/reasons-internal-external/.

Jollimore, Troy. 2005. "Why is Instrumental Rationality Rational? ." *Canadian Journal of Philosophy* 289-307.

Korsgaard, Christine. 1996. *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press.

Korsgaard, Christine. 1996. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, by Christine Korsgaard, 106132. Cambridge University Press.

Korsgaard, Christine. 2008. "The Normativity of Instrumental Reason." In T*he Constitution of Agency: Essays on Practical Reason and Moral Psychology*, by Christine Korsgaard. Oxford University Press.

—. 2014. The Sources of Normativity. Cambridge University Press.
McDowell, John. 1978. "Are Moral Requirements Hypothetical Imperatives?" *Aristotelian Society* Supplementary volume 13-42.
Nagel, Thomas. 2008. The Possibility of Altruism. Princeton University Press.

Parfit, Derek. 2011. *On What Matters.* Oxford University Press.
Sayre-McCord. 1986. "The Many Moral Realisms." Southern Journal of Philosophy 1-22.

Sayre-McCord, Geoffrey. 2007. *Essays on Moral Realism*. Cornell University Press.

Scanlon, Thomas. 2000. *What We Owe to Each Other.* The Belknap Press.

Setiya, Kieran. 2004. "Against Internalism." *Nous* 266-298.

Shafer-Landau, Russ. 2000. "A Defense of Motivational Externalism." *Philosophical Studies* 267-291.

—. 2003. *Moral Realism: a Defence*. Oxford: Oxford University Press.

Williams, Bernard. 1985. *Ethics and the Limitations of Philosophy.* Harvard University Press.