

Abstract: “Internalism, Practical Relations, and Psychologism”

Colin Patrick, Philosophy, University of Chicago

What exactly does it mean to say that a particular person “has reason to do” some particular thing? What makes a claim like this true or false? Whatever answer we give here, so it seems, will make a difference to what we take ourselves to be doing in giving advice to individuals, criticizing their actions and decisions, ascribing responsibility to them, and recruiting them for help in some cause, among much else. This is therefore not only an academic issue: it is not uncommon for activists, political leaders, public interest groups and others to make claims not merely about what “we,” collectively or as a society, have reason to do, but about what we have reason to do or avoid individually: for example, to adopt “greener” or more environmentally friendly consumer choices and habits, or to reduce our individual contribution to the negative environmental impact of industrial society more generally.

This issue of the reasons people have, and of what it ultimately means to attribute reasons to individuals, has been an especially pointed one in the English-speaking philosophical world over the last few decades in the wake of a well-known series of articles written by Bernard Williams (from 1980-2001) about the interpretation of statements of this form – that is, “N has reason to  $\phi$ ,” where “N” names a particular person and “ $\phi$ ” is an action verb. (I refer to such statements as “reason-attributions.”) Williams proposes two possibilities: “internalism,” where these statements are true only if  $\phi$ -ing serves some aim or practical interest N has, or connects up in some way to her “subjective motivational set,” or “S”; and “externalism,” where this is not necessarily the case, and where an individual can have reason to do something independently of her own aims – perhaps because she is morally required to, or because it would be for the best all around in some way, etc.

In these articles, Williams defends internalism as a coherent account of the

truth-conditions and normativity of reason-attributions, and, separately, argues against externalism. This dissertation in large part is meant to provide a limited defense of him on these points – though I depart in some important ways his conception of what internalism amounts to, and I disagree with the more standard view of what his main argument against externalism is.

His defense of internalism (which I discuss in Chapter 1) largely consists in distinguishing it from a sort of crude “sub-Humean” model, according to which an agent only has any reason to do what that agent, in that very moment, feels like doing or desires to do. Williams sketches out a more nuanced conception of an agent’s “S,” one that includes longer-term projects and goals, self-conceptions, “personal loyalties” and “sympathies,” among other things. On his account of internalism, therefore, one can have reason to do not only what will satisfy a present desire, but a range of things that are ancillary to one’s aims in a number of ways that may not in fact be grasped by the agent in question, and which may require reflection or deliberation to discover. It is thus possible, on his account, for an agent not only to have mistaken beliefs about her own reasons, but to go wrong with respect to them in practice – to act contrary to her internal reasons. This possibility, most will agree, is an important aspect of reasons as *normative* claims about what a person should (at least in some sense of “should”) do – one that a cruder account, based simply on what that person now feels like doing, would find it difficult, if not impossible, to account for.

Williams’s conception of what one’s “S” comprises is somewhat scattered and open-ended, which is a liability given that his internalism entails (somewhat controversially) that *all* reason-attributions have this as their basis. In Chapter 2 I try to shore up this conception as a basis for internal reason-attributions – without, I hope, oversimplifying it. I do this by drawing on the “threefold division of practical good” developed by Candace Vogler in her 2002 book *Reasonably Vicious*. Vogler identifies three ways in which an action (“A-ing”) can be explained as having a point: “use”

(where A-ing is a means to or part of B-ing), “fit” (where A-ing is fit to be done, not because it contributes to some finite task, B-ing, but because it is fitting or appropriate to a more standing commitment, e.g. a principle or rule to which she adheres), and “pleasure” (where A-ing is done because it is enjoyable or pleasant). Her interest there is in reasons that explain actions (or what are sometimes called “explanatory reasons”); I instead adopt this division as a basis for internal reason-attributions, i.e. a species of “normative reasons.” I argue that this division collects and sorts the various things Williams includes in his conception of an agent’s “S.”

It also, I think, helps to spell out the normative aspect of internal reasons – since each of these three dimensions of possible explanation can also count as a standard against which an agent can go wrong. (One can mistakenly do what is not a means, or what is an incomplete means, to B-ing; one can do what is inappropriate to a standing commitment one has; one can mistakenly do what is not in fact an example of the kinds of things one likes to do; etc.) The three forms above are able to make good on this because they each rest on what I call “practical relations” between an action, on the one hand, and, on the other, an aim to which it is an intelligible means, or a longer-term commitment to which it is appropriate, or a more general description of a range of actions (e.g. listening to early 20<sup>th</sup> century classical music) the agent enjoys doing and of which it is an instance. This sort of relation is interesting in its own right, I think, but also serves to underscore that the basis for internal reasons is not exhausted merely by references to the agent at all, much less to the agent’s “S.” One can go wrong by the lights of an internal reason, not simply because one can do what one no longer feels a desire to do, but because there are certain actions that do, and do not, intelligibly count as “practically related” to one’s aims, commitments, and enjoyments, quite independently of what one might think.

In Chapter 3 I discuss the second part of Williams’s argument about reason-attributions, i.e. his critique of externalism. Somewhat confusingly, he seems in these

articles to be putting forward two independent arguments, one of which, I think, is much stronger and difficult to rebut than the other (I'll refer to it as the "main argument") – while most of the critics of Williams focus on the weaker, secondary one. The main argument, so I maintain, involves an appeal to the fact that it is a distinctive form of proposition that is under discussion here – one that relates particular actions to particular persons – and not just any "general normative judgment" as Williams calls it, say a claim that  $\phi$ -ing is right or good or a claim that "there is reason" to  $\phi$ . (I sometimes refer to this point as the "distinctiveness condition" on reason-attributions.) Williams's argument seems to be that in moving away from consideration of the agent's "S," externalism detaches itself from anything that could pick out *this* particular agent as having reason to do *this* particular thing – and thus seems incapable of capturing what reason-attributions distinctively convey. Externalism seems, that is, to resolve reason-attributions into "something else, misleadingly expressed," as he puts it, or to make them replaceable by *other* distinctive claims, about what is right or good about the action, or about what "there are reasons," morally or prudentially speaking, to do. This is a problem for any theory that seeks to establish the truth-conditions of this particular form of statement.

This argument, which I explore in a number of places throughout, is largely overlooked in the critical response to Williams. Instead, most critics respond to a second, and I think independent, argument, to the effect that any putatively "external" reason, having no footing in the agent's "S," could not hope to move that agent to so act – whereas, Williams maintains, it is necessary, in order for a reason-attribution to be true, that the agent in question be able to appreciate (or come to appreciate) this reason well enough to act on it, such that the reason would then be the one that explains what she did or why she did it.

This last point is often referred to as the "explanatory condition" (EC) on reason-attributions. Many critics of Williams contend, against him, that considerations

quite “external” to an agent’s “S” *can* move agents to act, and that, therefore, they can pass the EC sufficiently well enough to count as that agent’s reasons, despite what Williams seems to be arguing. I argue that, even if it were true that some consideration could move an agent to act entirely independently of her prior motivations, this would not help clarify what exactly makes *that* consideration into *that* agent’s reason – as one can be moved to act by considerations that are not one’s own reasons in any clear, distinctive sense. (Much of the critical discussion of these articles resolves into a debate about whether practical deliberative conclusions can be reached independently of one’s prior motivations, and I argue here that this is a distraction from Williams’s main point, even though Williams himself is caught up in it.)

In Chapter 4 I briefly recap the more minimal account of internalism, based simply on the connections (or “practical relations”) between elements of an agent’s “S” and a certain range of actions, I develop in Chapter 2. I argue that, even though it is not (as Williams’s account is) committed to the EC, it is nevertheless sufficient to capture the sense in which reason-attributions are a) *practical* reasons, b) attributable in particular to the agents mentioned in them (i.e. it is sufficient to capture the distinctive content of reason-attributions), and c) not mere descriptions of the agent’s “S,” but count as claims about what, in *some* sense, it would be good, or sensible, for the agent to do. Many will likely want to reject the claim that one can “have reason,” even at this minimal level of normativity, to do something simply because it is “practically related” to an aim of theirs. No one, so the thought might go, can be conceived as having reason in *any* normative sense to do what is wicked or horrible. I counter by emphasizing that “having reason” does not, on this account, equate to “having *most* reason” (nor necessarily to having *moral* reasons), and it does not mean that what an agent thus has reason to do must be good or right. Internal reasons are about what it makes sense for one to do (and not to do) given what else one is up to, and my only point is that there is a distinctive, even if limited, conception of

normativity at work even here. It can, on this account, actually be a *bad* idea to encourage a person to do what they nevertheless have some reason to do; for the internalist, what a person has reason to do, and what it would be a good idea all-around to advise her to do, can come apart. I try to show that these points are not as threatening as they may seem – if only because, again, reason-attributions are only one particular kind of normative judgment that one can make about an agent and/or an action she might undertake.

One of the intuitions behind the EC seems to be that, when a reason that I have becomes one on which I in fact act, the very same reason that stood as a normative claim about what I should do can also explain what I did do and why, or reveal the point of what I did. Despite the fact that my minimal internalism does not entail that in order for a reason-attribution to be true, we have to have some kind of advance assurance that the agent in question is capable of appreciating this reason well enough to act on it, I argue in Chapter 5 that it *is* nevertheless sufficient to capture this more basic point Williams insists on about the potential symmetry between the normative and explanatory roles of reason-attributions. My thought here is that, once an agent actually *does* act on the internal reason, the same elements that spelled out how she had this reason in the first place (e.g. she aims to D, and  $\phi$ -ing is a means to that) can provide the basis for an explanation of the action given in terms of her reason for acting. Practical relations are in fact especially well-suited to this explanatory role: since they reveal some definite way in which it *would* be sensible for N to  $\phi$ , it is not surprising that they can reveal, once N acts, what the point, or the sense, of N's  $\phi$ -ing actually was. This may seem to reduce Williams's point about the explanatory power of reasons to a triviality ("the reason can explain the action insofar as the agent acts on it"), but, I argue further, not just any consideration an agent acts on (or which moves her to act) can rationally explain what she is doing. Such explanations, whatever else they may require, rely on at least some conception of what it *does* make sense to do

given what else (a particular aim, etc.), and this is what practical relations provide.

It may be thought that, since we very commonly appeal to reason-attributions in giving advice, and since the point of advice is to encourage a person to *act* in a well-advised way, and not simply adopt a correct belief about their reasons, then, contrary to what I am arguing, the EC must be a condition on reason-attributions: any true claim about what I have reason to do must be one on which I could be moved to act, on pain of rendering advice-giving pointless if not incoherent. Towards the end of Chapter 5 I counter this line of reasoning, maintaining that the *context* of advice-giving (where the point will perhaps be lost if I do not adjust what I advise the agent to do on the basis of what I can reasonably imagine her doing) is importantly distinct from its *content* – and that it is the latter that is under discussion here. (This is largely a repetition of the point that one’s internal reasons, on the one hand, and what it may be good all-around to advise an agent to do, on the other, can come apart for the internalist.)

In Chapter 6 I move beyond the claim that the EC is neither essential to a workable internalism nor to Williams’s main argument against externalism, and argue that it is moreover deeply problematic in itself: it is a condition on the truth of a reason-attribution that entails that the agent to whom it is attributed can appreciate, or come to appreciate, its truth (i.e. well enough to be prepared to act on it). This not only threatens circularity in the conception of their truth-conditions, but also seems to render even their limited normativity unintelligible: if I fail to appreciate some reason well enough to be able to act on it, while it is a condition on the truth of this reason-attribution that I *can* do this, will this mean that I am going wrong with respect to this reason in some way, or that I don’t so much as have it in the first place? On my more minimal conception, this problem does not arise. I can have reason to  $\phi$ , and go wrong precisely *because* I cannot appreciate this reason well enough to act on it. In order for a reason-attribution to stand as a normative claim, against which what I do (or fail to

do, for lack of an appreciation of its truth) can be seen as mistaken or insufficient in some way, so it seems, it cannot be a condition on its truth that I am able to appreciate it and act on it.

John McDowell suggests, in a well-known rebuttal to Williams, that the latter's internalism is "psychologistic." Insofar as the EC conflates the *truth* of this form of statement with its being *taken as true*, or believed – or makes the possibility of the latter into a condition on the former – this criticism, so I argue, is apt. This is not, however, McDowell's point: he accepts the EC as a condition on reason-attributions, merely disagreeing with Williams's claim that a process of rational deliberation, as opposed to something like "conversion," is how the externalist must meet it. (McDowell finds this restriction psychologistic because he thinks that deliberation is a process tied too closely to one's prior motivations to effect the "right kind of distance" between an agent's motivations and his reasons.) "Conversion," however, appears to be a process that is immediate and unarticulated, and thus in itself not, unlike deliberation, subject to any objective rational criteria. In accepting the EC, and then suggesting "conversion" as a possible means of meeting it, McDowell's proposed revision of Williams would therefore, I argue, leave it more psychologistic – at least as I understand "psychologism" – and not less.

Finally, in an ambitious and provocative book called *The Sources of Normativity*, Christine Korsgaard develops an account of the reasons individual agents have to do what they are morally obligated to do – one that arguably has the advantage, as internalism seems to have, of helping to sort out what makes a consideration *this* agent's reason in particular. It also incorporates Williams's idea (which I resist in the foregoing) that such reasons must be at least potentially motivationally efficacious for the agents to whom they are attributed. It stops short, however, of the internalist's claim that such reasons are (even in part) a function of the contingent configuration of the agent's "S." It thus may be thought to combine the advantages of internalism with

a disavowal of the latter's most troubling aspect – at least as far as moral reasons are concerned.

In Chapter 7 I undertake a close reading of the argument of this book, and argue that it fails to deliver these advantages. This comes in three stages. First, I examine what she calls the “normative question” – basically, a demand on the part of a reflective agent, placed in some particular deliberative context in which some moral consideration is failing to move her to act, to hear why she indeed should apply such a consideration to her case – and her criteria for any acceptable answer to it. Second, I apply this discussion to her analysis of the way in which David Hume, in passages from the *Treatise* and (especially) the second *Enquiry*, seeks, and, she thinks, ultimately fails, to answer it. I argue that, given the criteria she places on answers to the normative question, and her radically subjective conception of normativity (whereby a claim is normative only if the agent who asks the normative question *finds* it to be normative), she is ill-placed to be able to argue that Humean-style answers to the normative question *cannot* work. I argue that this renders her next move, to a Kant-inspired alternative, question-begging, and further – because her neo-Kantianism seems to entail that any reason I have must be a reason that *no one* could fail to find convincing – leaves her unable to account for the way in which reasons can be attributed differentially to particular individuals (i.e. to account for their distinctive content). Thirdly, I assess her neo-Kantian account independently of the charge of question-begging in the way she arrives at it. Here I argue that this account of our obligations and the reasons they give us winds up being self-defeating: it allows that an agent can fail to be moved by the account she offers of them well enough to ask the normative question with respect to them – and, by her own lights, when this happens it cannot suffice to simply reiterate that, and why, we have these obligations. Thus, the very reason she finds the Human answer to fail in answering the normative question will also show that her own account likewise fails. I argue further that, even though her

neo-Kantian account is, unlike mine, meant in part as an account of the normativity of moral considerations, it threatens much more directly than mine does the status of moral judgments as normative.