



Symposia Papers: Autonomist Internalism and the Justification of Morals

Stephen L. Darwall

Noûs, Vol. 24, No. 2, 1990 A.P.A. Central Division Meeting (Apr., 1990),
257-267.

Stable URL:

<http://links.jstor.org/sici?sici=0029-4624%28199004%2924%3A2%3C257%3ASPAIAT%3E2.0.CO%3B2-U>

Noûs is currently published by Blackwell Publishing.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/black.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Autonomist Internalism and the Justification of Morals

STEPHEN L. DARWALL
UNIVERSITY OF MICHIGAN

I. HOW SHOULD WE UNDERSTAND “WHY BE MORAL?”?

In its philosophically most interesting form, the question, “why be moral?” seeks a justification for morality. More accurately, it seeks a showing that morality’s demands themselves justify. It asks, why should I (or anyone) do what morality demands? What *reason* is there to do so? Of course, someone might ask this question as a simple challenge to point to something that will move him—something that will make him care. But while that must concern the moralist, and all of us concerned about morality, this challenge seems a philosophical threat only if it is relevant to whether morality can be justified.

There is a truistic sense in which it must be true that moral demands are justified. They are, we might say, morally justified. So it is tempting to reply simply, “because morality demands it,” and be done with it. But this response hardly satisfies. Without further comment, it is wholly internal to the system of reasons being questioned. It is no more satisfying a response to the philosophical request for justification than would “because etiquette requires it” be a satisfying response to the question of why we must abide by its demands.¹ In both instances, we want to ask whether considerations internal to a particular system provide any justification other than truistically, that is, other than within their respective viewpoints.

But by the same token, neither is “why be moral?” most interestingly viewed as a request for justification from within some *other* system of reasons or norms, say, those of prudence. Surely philosophers have attempted to justify morality in this way, but there

are two reasons why doing so engages a less interesting form of the question. First, even should it turn out that considerations of some other sort always dictate doing what morality demands, this would not constitute a justification of moral demands themselves.² Second, many of the same philosophical motives that prompt “why be moral?” should lead us to be similarly questioning of any other system of considerations and norms. Suppose we had a proof that moral demands always coincide with considerations of self-interest. Waiving the concern raised just, would this constitute a justification of morality? Obviously, it would do so only if considerations of self-interest are themselves justifying. Of course, they are truistically, from the point of view of prudence. But then moral demands themselves are already truistically justifying. When we ask “why be moral?” we do not put truistic justification in doubt; rather, we ask whether morality justifies in some way other than just from its own point of view. And in this same spirit we may also ask, “why be prudent?”

So what are we asking with “why be moral?” if we are asking for justification that is not merely internal to the moral point of view (so conceived), nor internal to the point of view of considerations of some *other* kind? The answer, I think, is that we must be asking whether considerations that present themselves as reasons from within the moral point of view, indeed as uniquely overriding reasons, *really are reasons*. In the same spirit, we might ask whether considerations of prudence really are reasons. In both cases, a familiar way of thinking exists in which certain considerations are treated as providing reasons to act. In asking the question, we step back from that way of thinking and ask whether the considerations treated as reasons within it are so in fact. We take the consideration with the qualifier—moral reasons or prudential reasons—and ask whether they are reasons, as it were, with the qualifier removed. As I shall put it, we ask whether they are *unqualified reasons*, or reasons *unqualified*, though by that we do not necessarily mean either unconditionally or overridingly. We simply ask whether they are reasons *sans phrase*.³ Actually, since morality claims decisive authority for its demands, there are really two distinct questions in its case. First, do moral demands have any unqualified weight as reasons at all? And second, if they do, is their weight always overriding?

But what could it mean for something *really* to be a reason, and not just from within a given perspective? A reason for *acting* must be something an agent can *have*—something she can grasp, indeed grasp *as a reason* and act on. So the contrast between reasons from within a specific viewpoint and unqualified reasons cannot be a contrast between something perspectival and something which isn’t—

like that between apparent and real shape. Considerations from within a point of view are reasons unqualifiedly only if the point of view itself makes a normative claim on an agent—if it is one by whose considerations the agent ought to be guided. And, to anticipate a bit, a rational person will only be able to grasp a consideration's weight as an unqualified reason if she has available some standpoint other than the qualified point of view (so conceived) from which she can appreciate this normative claim—from which she can judge that she ought to be guided, to some extent or other, by considerations from the more specific point of view.

II. DEFLATIONIST THEORIES OF JUSTIFICATION

Objections of two sorts may be raised to the dialectic thus far. First, it may be held that justification relative to a point of view is the only kind of justification there is. If this is so, then any search for unqualified reasons will be fruitless; there are none. Relativism of this sort has not, however, proven a very attractive philosophical position. A second objection is apt to be more appealing. It holds that, while there is such a thing as unqualified justification, that is not because some further normative grounding exists for the point of view relative to which unqualified reasons are justifiers. Rather, there is a perspective (usually some version of the agent's own standpoint) which is justificationaly privileged in the sense that being unqualifiedly justifying *is the very same thing* as being justified from that point of view. Thus, it may be held, being justified in view of the agent's overall interests, or her present ends, or her informed desires is what being unqualifiedly justified is.

We can call views of this second sort *deflationist* theories of unqualified reasons or justification. They aim to show how some considerations can make an unqualified normative claim on the agent from *within* his own point of view without supposing any normative backing for the perspective itself. For example, a deflationist instrumentalism would hold that being the best means to the agent's ends is unqualifiedly justifying, not because he stands under a universal norm prescribing instrumentally rational conduct, but because an act's being unqualifiedly justified just is its being instrumentally rational. Likewise, informed desire theories of rational justification may be proposed with deflationary intent. Here again, the idea is that while the agent's informed desires make a normative claim on him and thus create unqualified reasons to act, that is not because as rational he is bound by a norm prescribing that agents do what will achieve their informed desires. Rather, unqualified justification and its normativity are nothing over and above bearing the right relation to informed desire.

Now justification must indeed run out at some point, but it may be wondered whether deflationist theories do not make it run out too quickly. According to a deflationist informed desire account, for example, the question why we should do what we informedly desire cannot even arise.⁴ That would be like questioning why we should do what is unqualifiedly justified, and surely that is senseless. But the first question needn't seem as senseless to us as the latter. I know I informedly want something, but that you do too, and we cannot both have it. Can I not naturally ask whether I am (or any person would be) justified in taking it? If I can sensibly ask what the best reasons recommend in such a case, I can do so only by assuming that being justified is not the same thing as being what would realize my informed desire.

But what if deflationism were true? What if being unqualifiedly justified just were being coherent with the agent's ends or, more plausibly, with her informed desires. How then would the justification of morality stand? By that I mean less how would morality fare on that understanding of what it is to be justified; rather, my question is: how would we regard its faring however it might on that understanding?

Suppose we become convinced that an act's being unqualifiedly justified is the very same thing as its being coherent with the agent's present ends. And suppose that what morality demands is frequently not what would cohere most with these. How then should we regard the fact that what morality demands is not unqualifiedly justified on our present understanding of what that is? It is hard to see why it should pose any new threat at all. We will have presumably already known that what morality demands is not always most coherent with the ends agents have. That has never been in dispute. What we will now be convinced of is that in such situations agents' flouting moral demands is unqualifiedly justified. But that does not mean that these agents have any *further* justification, other than whatever justification we have always thought provided by coherence, for flouting morality. It will not mean that they stand under any norm of practical reason counselling them to act most coherently with their ends, and thus counter-morally. Moreover, we will realize that as far as any norm implicit in coherence itself goes, agents would be no less justified in simply renouncing offensive ends, and adopting those more congenial to moral demands.

Or again, suppose that we become convinced that unqualified justification is the very same thing as being what an agent would be dominantly moved to do by his informed desires? We will presumably have always known that the informed desires of some agents on some occasions would lead them to flout moral demands.

That has only rarely been in dispute. What we now become convinced of is that this flouting is unqualifiedly justified. But we do not become convinced thereby that this has any justification we ever doubted it had. Our conviction has no additional normative force beyond what we always were willing to think. In both cases what will have happened is not that we will have come to believe that considerations provided by the agent's own ends, or through his informed desires, have more normative force than we thought. Rather, what will have happened is that we will have come to think that unqualified justification has *less* normative force than we thought.

If this were so, it seems likely we would feel less philosophical pressure to show that moral demands are unqualifiedly justifying.⁵ If that is all unqualified justification is, we might think, why does morality stand in need of its support? There is one reason why we might still think it does. The reason we seek a showing that moral demands are unqualifiedly justified is not just to show that people never have weightier unqualified (and fully normative) reason for acting otherwise. If deflationism is the truth about justification, then this will be true by default, since there is no such thing as the fully normative justification we had supposed there to be. Still, our unusual understanding of moral demands includes the positive idea that they *do* provide fully normative justification.

Now obviously, if deflationism is true, this must be false. Since I know no way of proving deflationism false, or of proving that there are fully normative, unqualified reasons, I suggest the following strategy. I propose that we help ourselves to the assumption that it is false, and consider the question of whether moral demands are unqualified justifiers on that assumption. After all, the assumption that fully normative unqualified reasons exist does not presuppose in any obvious way that moral demands are among them. We could then ask this question: assuming there is such a thing as unqualified justification with full normative force, are moral demands unqualifiedly (and overridingly) justifying in this way.⁶

In the space remaining, I want to indicate a strategy for approaching this question. I aim to do so with another question: supposing there is such a thing as unqualified justification, in what should we take it to consist? I hope to make it plausible that on one attractive theory of what unqualified justification would consist in, moral demands (properly understood) might indeed be unqualifiedly justifying.

III. INTERNALISM AND THE NORMATIVE

Internalism as a doctrine about reasons has enjoyed significant recent support partly because it provides the hope of capturing the

idea of normative force consistently with a broadly naturalistic philosophy. By internalism, I mean the doctrine I have elsewhere called *existence internalism*, rather than what I there called *judgment internalism*.⁷ Judgment internalism about reasons is the view that a *judgment* that a reason exists to do something entails some motivation in the judge to act accordingly. Existence internalism, on the other hand, holds that a necessary condition of *there being* a reason for someone to do something is his having a motive, although not necessarily under the circumstances he is presently in. Among existence internalisms are those we might call *constitutive internalisms*. They hold that the motivational condition is not merely necessary; it is what being a reason consists in.

The deflationist informed desire view is a constitutive internalism, since it holds that an act's being rational just is its being what the agent would be dominantly motivated to do on full and vivid consideration. And, as a deflationist view, it holds that the normative force of reasons is fully constituted by the motivational pull a consideration exerts when considered in light of knowledge and experience. Suppose we think of constitutive internalisms as having the following general form:

p is a reason for S to do A if, and only if, were S to consider p in the right way he would be given some motivation to do A.

This general formulation does not necessarily aim to reduce the normative force of reasons to the motivation a consideration itself provides, in the way a deflationist internalism would. It retains a normative notion in the idea that reasons are considerations that motivate when considered in the right way—if you like, as an agent *ought* to consider them.

Now even the informed desire view wants to retain the distinction between *de facto* motivation and a reason or justification. It does this by distinguishing between what actually motivates and what would motivate if the agent were informed, experienced, and vividly imagining. But it does not conceive of this latter condition in normative terms. It assumes no burden to argue that this is the way in which an agent *ought* to consider candidate reasons, or, *a fortiori*, that it is the only way. As a deflationist theory, it rejects any need of such a premise. A reason just is a fact which would motivate when considered in this way; its normative force is simply the pull felt from this point of view.

But, as we saw, we may also want to ask, “should I do what I would want most were I to consider things in this way?” We feel this to be a genuine question, and, we may think, if there is such a thing as fully unqualified justification, there should be an

answer to it. So we cannot simply accept that candidate reasons, which motivate on a given way of considering them, are unqualifiedly justifying—or, *a fortiori*, that they provide the *only* unqualified reasons—without some reason to think that the way of considering is itself normatively grounded. We need to think that it is a way in which we ought to be considering things.

IV. AUTONOMIST INTERNALISM

Although philosophers are perhaps most familiar with internalism in its deflationist versions, driven primarily by philosophical naturalism, historically there has been another significant source of support for the idea that the normative must be located within the will. This is a position that may be called *autonomist internalism*. With roots in Stoicism, but beginning in the modern period in the late seventeenth century, and running through Rousseau and Kant, this tradition has maintained that a free rational agent can only be bound by constraints emanating from his own will.⁸ Norms and justification cannot stand independent of the will, simply confronting it as external constraint; rather, their normativity derives from self-imposition in the activity of autonomous living. Autonomist internalists take it that there is a form of practical thinking which makes autonomy possible, and that unqualified justification derives from motivation arising when this form of thinking is properly engaged. Propositions about what a person would be justified in doing, and hence what he ought to do, are made true by being realized in *autonomous* will.

But how are we to conceive of autonomy itself? The root notion that emerges in the modern period is that autonomy consists in the self-critical questioning of standards by which to live in search of unqualifiedly justifying ones and the commitment to live by these. About this there is general agreement: autonomy requires a capacity to step back from current motives and beliefs about justification, to evaluate these self-critically, and to guide oneself by motives and standards one can self-critically and unqualifiedly endorse.

Now a kind of circularity seems involved here, and it is. Unqualified justification is what would guide the autonomous agent. And autonomy itself involves the search for what one can regard as unqualifiedly justifying. It is no accident that the increasing tendency in the modern period to think of fundamental questions of justification as “autonomous” and distinct from, although perhaps related to, questions of metaphysics, psychology, theology, custom, and so on, goes hand in hand with the development of conceptions of the autonomous agent. The very raising of fundamental ques-

tions of unqualified justification in this way is itself an exercise of autonomy. What autonomist internalism asserts is that answers to questions raised through this exercise of autonomy are themselves realized through autonomous practical thinking.

Unlike deflationist internalisms, autonomist internalism retains a normative element in its ideal of autonomy. And this is just what we should expect. We saw in Part II that it seems impossible to capture the full normativity of unqualified justification which autonomous practical activity seeks without retaining a normative element at this point. But how is this ideal to be specified?

Locke described a species of autonomy that is realized in suspending present desires, examining their objects “on all sides” and being led thereby to a true view of one’s greatest *pleasure* (which he believed to be any rational agent’s final end) in a way that raises the dominant desire to seek it.⁹ Shaftesbury, while no hedonist, also believed that a person can only be the “author” of his own life by engaging in a “self-converse” that enables him to align present desire with his own greatest *interest*.¹⁰ More recently, W.D. Falk has defended an informed desire account of justification motivated by autonomist internalism and by an ideal of the autonomous agent as one who lives nonevasively by making motives responsible to reality as he sees it.¹¹ These form an ordering of ideals of autonomy. Locke’s agent can step back from present desires, but has no standpoint to raise the question, “why should I do what will give me most pleasure?” Shaftesbury’s agent can sensibly ask this question, but not “why should I do what is in my interest?” Falk’s agent can ask that, but cannot sensibly raise, “why should I do what I will want most when fully informed and mindful of what I know?”¹²

In Part II, I suggested that the search for unqualified justification drives us to raise questions at each of these levels. It would seem, therefore, that for an autonomist internalism adequately to capture the idea of unqualified justification, it will require a less restricted ideal of autonomy. The autonomous agent must be able to raise all three of these questions for herself. Now we have an explanation for why deflationist accounts fail to give unqualified justification the unrestricted normativity we seek. The search for unqualified justification is itself an exercise of autonomy through which we step back, not just from our present ends, but also from ends we continue to have through knowledge and experience. What we seek is something that can have a grip on us, not *simply* within the standpoints provided by our ends, or our informed desires, or the various other standpoints we share, through identification, with others—friends, family, community, and so on. What we seek is something that can grip us *as someone who can raise the very question we are raising*.

And this suggests a standpoint from which we can grasp our question. We wonder what grips us from the standpoint of an agent driven to raise our very question. And autonomist internalism should answer: what *we* would grip from that standpoint. Since we seek norms to guide any such agent, we must ask: what norms would we will for all from that point of view? But how do we interpret that? Here I can give only the briefest sketch, but you will recognize something of its general shape. Since we want to know what we would will for all from the perspective of an autonomous agent as such, we might conceive of the relevant standpoint as one that is impartial between such agents, with a veil of ignorance regarding more specific features and ends, but with full general information otherwise. As a basis for this choice we assume we have a governing interest in a rationally autonomous life, and that this involves, albeit *defeasibly*, an interest in pursuing ends and projects that arise from our own, yet critically informed, standpoints, whatever these might be. We also suppose that we share a capacity and desire to constrain our pursuit of individual ends, and of group projects with which we identify, by principles which we would will for all from this (impartial) point of view.¹³ We then interpret “What principles would we will or legislate for all from this standpoint?” as “What principles would it be instrumentally rational for (any of) us to choose, from behind a veil of ignorance, motivated by the assumed ends, to guide the conduct and practical thinking of all (similarly capable) agents?”

It seems clear that, for circumstances anything like ours, it would be rational to choose that agents regard themselves as *presumptively* justified in pursuing projects, individual and group, with which they can critically identify from their own points of view (including as these are expanded through shared connections to others). But by the same token, again, for circumstances anything like ours, it seems clear that it would be rational to choose that agents regard themselves as bound by principles constraining these pursuits, which are otherwise *presumptively* justified, in the common interest. Whatever the details of such principles, it seems plausible that it would be rational to choose that agents, in circumstances like ours, take themselves to be constrained by some conception of moral demand (as we might call it). This would mean that, not only do moral demands (suitably conceived) give us unqualified reasons; they give us overriding unqualified reasons.

A moment to retrace our steps. I first argued that “why be moral?” is best understood as a question about unqualified justification, and, next, that the leading, deflationist, theories of justification are implausible as accounts of unqualified justification, but, then, that even if we were to come to be convinced they are cor-

rect, that would pose no new philosophical threat to morality. So I then assumed that deflationist theories are false, and that we should take it that there is such a thing as fully unqualified justification. My question then was, what can we reasonably understand unqualified justification to be, and, do moral demands provide unqualified reasons on this understanding? My answer has been that unqualified justification is best understood through autonomist internalism. The very search for unqualified justification is an exercise in autonomy; and the only thing unqualified justification is likely to be is something we can construct in the course of that search. Finally, I have claimed that the conception of unqualified justification we will then have constructed will be one in which moral demands do indeed provide unqualified reasons.

NOTES

¹The example comes, of course, from Philippa Foot, "Morality As a System of Hypothetical Imperatives," *Philosophical Review* 81 (1972), 305-316.

²I take this to be the thrust of H.A. Prichard's "Does Moral Philosophy Rest on a Mistake?," *Mind* 21 (1912), 21-37.

Dan Brock argues that, because moral demands claim superiority over other reasons, and because the only possible (nontruistic) justification would be in terms of some other system of reasons, morality cannot possibly be justified. He does not find this disturbing, because he thinks we do not regard our willingness to embrace other systems of reasons to depend on their being able to be justified externally. This seems right if external justification is taken to mean "in terms of another system of reasons for acting." However, I doubt that this is the only way in which a (proposed) system of reasons can be (nontruistically) justified. "The Justification of Morality," *American Philosophical Quarterly* 14 (1977), 71-78.

³This formulation is unhappy in one way since W.D. Ross used the formulation 'duty sans phrase' to contrast with 'prima facie duty'. But our contrast is not with *prima facie* reasons, since *prima facie* reasons really are reasons, albeit other things equal; it is with something like "apparent reasons", or "proposed reasons". See W.D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 18f.

⁴Or, at least, it cannot arise as a *practical* question. There is still the theoretical question, why is it that being unqualifiedly justified is the same property as being what would realize informed desires.

⁵Thus philosophers like Foot, Brandt, Railton, and others, who are deflationists, in my sense, of one stripe or another, are not likely to think that moral demands stand in need of being shown to provide unqualified reasons to act.

⁶For ease of exposition, from here on I will use 'unqualified reason' and 'unqualified justification' to include full normative force.

⁷See *Impartial Reason*, p. 54. Two classic articles on the distinction between internalism and externalism in ethics are W.D. Falk, "'Ought' and Motivation," *Proceedings of the Aristotelian Society* 48 (1947-48), 111-139; and William Frankena, "Obligation and Motivation in Recent Moral Philosophy," in A. Melden, eds., *Essays in Moral Philosophy* (Seattle: University of Washington Press, 1958). A recent discussion is Christine Korsgaard, "Skepticism About Practical Reason," *Journal of Philosophy* (1986).

⁸For some discussion of this tradition in late seventeenth century England, especially in the writings of Shaftesbury, see my "Motive and Obligation in the British Moralists," in Ellen F. Paul, ed., *The Foundations of Moral and Political Thought* (Oxford: Blackwells, forthcoming).

⁹Locke, *An Essay Concerning Human Understanding*, ed. with a foreword by P. H. Nidditch (Oxford: Clarendon Press, 1985), p. 263.

¹⁰Especially in "Soliloquy, or Advice to an Author," in *Characteristics of Men, Manners, Opinions, Times* [1714], ed. John M. Robertson, intro. by Stanley Grean (Indianapolis: Bobbs-Merrill, Co., Inc., 1964).

¹¹Falk has worked out this view in a number of papers collected in *Ought, Reasons, and Morality* (Ithaca and London: Cornell University Press, 1986). These include "On Learning about Reasons," "Action-Guiding Reasons," and "Morality, Self, and Others." Crucially, the view requires not just that the agent discover the truth, to which his motives must be made responsible, but that he *mind* it appropriately.

¹²Some care is necessary here. Plainly each agent may ask, "why is it the case that I should . . . (i.e., why is this unqualifiedly justifying?)" But here they will be asking a purely theoretical question about what justification is. What they cannot do is ask this as a practical question. This puts pressure on their views as autonomist internalisms, since I take it that autonomist internalism aims to blur any sharp distinction between true theoretical claims about what justifications is and practical norms of justification (i.e., between "meta-justification" and "normative justification").

¹³That is, that we have a capacity and desire to be *reasonable*, in Rawls's sense. Needless to say, many of the aspects of this sketch are adapted from Rawls's writings. In addition to *A Theory of Justice*, see esp. "Kantian Constructivism in Moral Theory," *The Journal of Philosophy*, 77 (1980), 515-572.