

Internalism about reasons: sad but true?

Kate Manne

Published online: 4 December 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Internalists about reasons following Bernard Williams claim that an agent's normative reasons for action are constrained in some interesting way by her desires or motivations. In this paper, I offer a new argument for such a position—although one that resonates, I believe, with certain key elements of Williams' original view. I initially draw on P.F. Strawson's famous distinction between the interpersonal and the objective stances that we can take to other people, from the second-person point of view. I suggest that we should accept Strawson's contention that the activity of reasoning with someone about what she ought to do naturally belongs to the interpersonal mode of interaction. I also suggest that reasons for an agent to perform some action are considerations which would be apt to be cited in favor of that action, within an idealized version of this advisory social practice. I then go on to argue that one would take leave of the interpersonal stance towards someone—thus crossing the line, so to speak—in suggesting that she do something one knows she wouldn't *want* to do, even following an exhaustive attempt to hash it out with her. An internalist necessity constraint on reasons is defended on this basis.

Keywords internalism about reasons · normative reasons · motivations · practical reasoning · the interpersonal stance · Bernard Williams · P.F. Strawson

K. Manne (✉)
Department of Philosophy, Cornell University, 225 Goldwin Smith Hall, Ithaca, NY 14850, USA
e-mail: kam468@cornell.edu

Internalists about reasons (or, equivalently, reasons internalists) endorse a claim to the effect that an agent's normative reasons for action are constrained in some interesting way by her desires or motivations.¹ The classic version of this position was first defended by Bernard Williams, in his "Internal and External Reasons" (1981a, originally published in 1979). Williams initially introduces the internalist on the scene as someone who believes that sentences of the form 'A has a reason to ϕ ' "imply, very roughly, that A has some motive which will be served or furthered by his ϕ -ing, and if this turns out not to be so the sentence is false;" whereas, according to the externalist, "there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive." (1981a, p. 101) And Williams subsequently clarified his own preferred version of the position as being the claim that "[an agent] A has a reason to ϕ only if he could reach the conclusion to ϕ by a sound deliberative route from the motivations he already has." (1995a, p. 35) Now, as we are going to see, how to envisage the process of sound deliberation here is a live and delicate issue. But, for the moment, we can imagine A coming to learn various non-normative facts which bear on his situation, and having procedural errors in his reasoning corrected, among other things. If these and other such sound deliberative processes would not induce A to form the 'conclusion' to ϕ —or, most likely better, to form a *motivation* to ϕ which might be overridden and thus fall short of a conclusion—then this agent will have no reason to go around ϕ -ing, according to the internalist.² This is so even if others will be made miserable if he doesn't.

No prizes, then, for guessing that internalism about reasons has been highly controversial. In particular, it has been held to be a claim with an obscure rationale and deeply counterintuitive implications.³ But I'm inclined to believe otherwise. I believe that the claim has a relatively clear, if controversial, rationale and that it may represent a deep, if sad, truth about practical normativity and our moral-cum-social relations. So here, I will wind up defending a version of reasons internalism—and one which comes fairly close to Williams' own position, but is in some respects even stronger than the necessity condition he proposed.

My aim in this paper is twofold, however. I both want to offer a new argument for internalism about reasons—although one that resonates, I believe, with certain key elements of Williams' original view—but also to explore a certain perspective

¹ A few quick pieces of housekeeping: when I talk about reasons for action throughout this paper, I will always mean *normative* reasons for action—i.e., roughly, considerations which go some way towards justifying an action—rather than so-called *motivating* reasons—i.e., considerations which might be held to be the contents of the thoughts that dispose an agent to act on a particular occasion. And I will be concerning myself exclusively with reasons for action proper, partly on the grounds that reasons internalists have historically taken their thesis to apply solely to such reasons, rather than (e.g.) putative practical reasons to want certain things or be in certain emotional states. Reasons for action are certainly the focus in Bernard Williams' original discussion (1981a). But they remain the focus in the careful and state-of-the-art discussion by Stephen Finlay and Mark Schroeder, who characterize 'schematic internalism' as the claim that "every reason for action must bear relation *R* to motivational fact *M*." (2012, p. 3)

² Note that Williams sometimes adopts the simplifying assumption that reasons are 'conclusive' or 'all-in' reasons. (See, e.g., 2001, p. 91.) My revision here is intended to allow this assumption to be dropped.

³ See, e.g., Derek Parfit, who has recently declared Williams' views here to be "baffling" (2011, vol. 2, p. 435)—with real pain, given Parfit's evident deep respect for Williams. I'll call on Parfit to act as a foil for Williams in several places in what follows.

on practical normativity from which this argument makes sense. This perspective is what I'll call a practice-based approach here, admittedly rather blandly.⁴

My starting place in these endeavors will be P.F. Strawson's influential distinction between the interpersonal and the objective stances that we can take to other people, from the second-person point of view (1962/2008). I suggest that it is only when we relate to other people as such, thus adopting the interpersonal stance towards them, that we can be said to reason *with* them. This is as opposed to ordering them about, coercing them, or trying to 'manage' their behavior (among myriad other possibilities). I then suggest that reasons are the kinds of considerations which would ideally be apt to offer to another person when we are reasoning with her, or (similarly) offering her collaborative advice or friendly suggestions, about what she ought to do. Somewhat more precisely, reasons for an agent to perform some action should be understood as being considerations which would be apt to be cited in favor of her performing it, when she is being reasoned with by someone ideally suited to play this social role—i.e., a well-informed and well-disposed person who constitutes her ideal advisor. It follows that reasons for an agent to perform some action are considerations which would be apt to be cited in favor of that action, within an instance of the interpersonal activity of reasoning with the agent. This is the intended upshot of Sect. 1 of this paper.

I then suggest, in Sect. 2, that there's a deep and confronting claim which we should acknowledge as at least plausible. Namely, there may be real limitations on what can be said and done, which we might have hoped could be said and done, within the interpersonal activity of reasoning with an agent. Take Williams' example of a man who is nasty to his wife (1995a). And suppose—admittedly with some hubris—that we ourselves are playing the part of his ideal advisor. It seems that there might be real limits as to what we can aptly say to him within the confines of this practice, to get him to change his ways. The internalist insight can be construed roughly like this, I propose: if this man would be totally unmotivated to treat his wife more nicely, even following an exhaustive attempt to hash it out with him, then we cannot aptly suggest to him in the interpersonal mode that he be nicer to his wife. His lacking the relevant sort of motivational propensity means that saying as much would no longer have the status of a suggestion. The point has been rendered moot; the conversation has gone dead. We may have to retreat to the objective mode instead then, and try to manage his behavior and the situation—e.g., by marching him off to therapy, or helping his wife to leave him, or trying to get him arrested (if he is positively abusive). Or we may have to retreat *simpliciter*—i.e., simply walk away from the situation entirely. And there are other, subtler possibilities as well here, which I'll go on to consider later.

Putting the key ideas of Sects. 1 and 2 together though, it follows that the callous husband does not have a *reason* to treat his wife more nicely. For, insofar as he lacks a suitable motivational propensity to do so, we cannot aptly suggest to him that he be nicer to his wife in the interpersonal mode, from which it follows that there is

⁴ 'Pragmatism' or 'humanism' might be more descriptive labels, but they are also potentially more misleading, given at least some of their respective connotations. Thanks to Nicholas Smyth for useful discussion on this point.

literally no reasoning with him about this particular matter. Hence, the relevant moral reasons claim in turn stands defeated, in lacking this crucial connection with the agent's motivations.⁵ However, and as I'll discuss in Sect. 3, we can—and I think often should—make the following kinds of moral claims instead: this man ought to be the kind of person who has a reason to treat his wife more nicely. Indeed, he may be subject to certain criticisms precisely *because* he lacks this reason.⁶ It is also simply bad that the man treats his wife in this way. And we ourselves may have reasons to try to change this man, or at least to try to prevent him from continuing to mistreat his wife. All of these claims are perfectly consistent with reasons internalism, and we can continue to make them, all the while acknowledging what I take to be the internalist's central insight. Namely, there is an important and in my view distinctive kind of normative claim—e.g., that this man has a reason to do otherwise—which we are sometimes forced to give up, owing to deficiencies in the agent's motivational profile. This is sad but likely true.

1 Reasons in practice

It's not hard to see why would-be defenders of reasons internalism often provoke an incredulous, or simply horrified, reaction. Suppose one starts out by saying, as many contemporary theorists do, that practical reasons are just the basic unit of practical normativity, the moral philosopher's equivalent of a sub-atomic particle.⁷ Then one announces—perhaps apologetically, perhaps blithely—that the putative reason for an agent A to ϕ can hold only if A has some sort of motivation that would be served

⁵ Note that, although internalism about reasons is sometimes billed as a thesis concerning desires specifically, I prefer to put things in terms of motivations to act, in keeping with the spirit of Williams' original discussion. For, as we have seen, Williams initially formulated his claim in terms of 'motives' rather than desires, and went on to characterize the 'formal' notion of a desire he was working with in such a liberal way as to rule out very little by way of behavioral dispositions of a broadly conative nature. He included in this category "such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent." (1981a, p. 105) I want to at least leave room to be similarly inclusive, and some theorists would understandably balk at construing the notion of desire so broadly. Thanks to Amartya Sen for helping me to see this, and also for illuminating discussions of Williams' views generally.

⁶ Why focus on moral reasons and their surrogates here, specifically? The argument I'll develop for reasons internalism is intended and formulated so as to apply to reasons for action across the board. But I focus on moral reasons throughout this paper because the internalist conclusion seems especially worrisome as applied to them in particular.

⁷ This is a metaphor that is once mooted by Parfit, when he writes that, in order to see what various first-order normative theories and principles imply, "we must answer questions about reasons. That is like the way in which, to know about the nature and properties of atoms, we must answer questions about sub-atomic particles." (2011, vol. 1, p. 149) But since Parfit does not believe in fundamentally different kinds of practical reasons, it might be more accurate to say that they are supposed to be the equivalent of the one true sub-atomic particle. For another relevant metaphor here, see Schroeder's reference to reasons as being (presumably, discrete) pros and cons on a list in the mind of God (2007, p. 166).

or promoted by her ϕ -ing.⁸ If I said this, then I think it would be natural to look at me pretty darkly. “Can’t people lack all sorts of crucial motivations?” you might very reasonably say. For example, Harry may be unmotivated to help in the fight against global warming. So what if he is unmotivated? Surely we may still be entitled to criticize him and urge him to do better. Indeed, doing so might be vital and hence positively obligatory in cases with this structure—i.e., collective action problems.⁹

All fair enough, to my mind.¹⁰ But I’m inclined to reject the opening move here: namely, where we say, with Derek Parfit along with T.M. Scanlon (1998), and many other contemporary theorists besides which, that practical reasons are just the basic unit of practical normativity.¹¹ For, I’m not convinced that there is anything like a sub-atomic particle when it comes to practical normativity. I’m not convinced that practical normativity works in such a way. Of course, this is not to deny that there are real attractions and also benefits of taking it to work in this way (insofar as in calling it an ‘it,’ we are not already headed towards monism if not atomism). Still, the idea that there is one basic unit of normative measurement, which can encompass the whole domain without distortion, strikes me as just one way of

⁸ I take it that internalists need not endorse the converse claim, and very well may not. Williams wrote that “the internalist view of reasons for action is that this formulation provides at least a necessary condition of its being true that A has a reason to ϕ ... It is a further question whether the formulation provides a sufficient condition of an agent’s having a reason to ϕ .” (1995a, pp. 35–36) He went on to say that he *did* think it was probably a sufficient condition as well, but that this was a separate issue, and one he wouldn’t take up in the context of discussing reasons internalism. And he never defended this claim at any length, at least to the best of my knowledge. I’ll adduce my own reasons for positively doubting it later on.

⁹ It should not be overlooked that Williams’ original discussion of reasons internalism ends with an application to the issue of public goods and free riders, which is held to lie “very close to the present subject.” (1981a, p. 111) For, as we will go on to see, Williams’ internalist thesis effectively suggests that there may be a gap—perhaps even a vast gap—between what it is good or desirable that people might do collectively, or what it is otherwise good or desirable to have happen, versus what individual people can actually be expected to do insofar as they are behaving reasonably.

¹⁰ However, some theorists would want to resist the idea that Harry could lack a motivation to do the morally right thing, at least under conditions of full procedural rationality. See Julia Markovits (2014) for a defense of such a view, and also for an interesting and novel argument for internalism about reasons quite different from my own here. Schroeder has defended a somewhat similar, broadly ‘Kantian’ view, albeit via a very different route. Schroeder suggests that moral reasons may be massively overdetermined, such that more or less any desire on the part of an agent would be promoted by doing what it is morally right to do. (2007, §6.3) I discuss Schroeder’s ‘Hypotheticalist’ version of the Humean Theory of Reasons in detail in work in progress.

¹¹ The idea that reasons are normatively ‘basic’ or ‘fundamental’ has been becoming increasingly popular, and manifests itself theoretically in a variety of claims. Particularly relevant for my purposes here is Parfit’s claim that the concept of a reason is fundamental in the sense that something matters only if we have reasons to care about it (2011, vol. 1, p. 148), and also that someone who is cognizant of the relevant ‘reason-providing facts’ can be criticized for acting in some way only if they have reasons to conduct themselves differently (2011, vol. 2, p. 442). Parfit holds, more generally, that normativity always “involves reasons or apparent reasons.” (2011, vol. 1, p. 144) Note too that the idea that reasons are normatively basic cuts across naturalist/non-naturalist and reductionist/non-reductionist party lines. Schroeder, a card-carrying reductive naturalist, endorses ‘Reason Basicness’: the claim that “what it is to be normative is to be analyzed in terms of reasons,” (2007, p. 81) while also thinking that the property of being a reason for an agent can be reduced to the property of being desired by her.

looking at things—one picture among many. It seems to me worth exploring other, rival pictures too.

Namely, when it comes to understanding the domain of what we have taken (fairly recently) to calling ‘practical normativity,’ I’m attracted to the idea of beginning with an account of human practices and activities. Forget the concept of a reason, or any other normative concept, for a moment. Think first instead—I’d propose, as a practice-based theorist—about what we *do*. Think about our practices of talking to each other, and reasoning with each other, as well as by ourselves. Think about more than that, too, though: think about the ways we instruct, reproach, request, cajole, wheedle, manipulate, demand, condemn, yell, and even stamp our feet on the ground in disgust at people’s conduct. Think, in other words, about the whole teeming mess of embodied and socially-situated normative *behavior*—i.e., behavior by means of which we give voice to ideas about what to do, and also what should happen. We might hope to specify the job description (as it were) of various abstract normative notions in terms of the *role* which they play in human practices of this kind. And, once one adopts such a practice-based perspective, it would actually be a bit surprising if one all-purpose normative notion—the notion of a practical reason—could work as the contributory notion vis-à-vis the *telos* of all of these different critical practices in which we engage with other people. A certain sort of pluralism naturally becomes the default. Thus, from a practice-based perspective, the subject matter of moral philosophy may be more akin to biology than to particle physics. And the sub-atomic particle picture may have something like the status of the longstanding myth of vitalism.¹² For, just as there is no *élan vital* differentiating living organisms from non-living things, there may be no one distinguishing mark of practical normativity. Practical normativity may be more of a hodgepodge of various critical and other related games we play, which are deeply interwoven but irreducibly diverse. (How diverse being a question for ongoing investigation.)

You might think, at this point, that this sounds suspiciously Wittgensteinian. It’s an association which I personally wouldn’t disavow. Admittedly, you might feel differently on this score though. But bear with me: for while it’s hard to give a non-question-begging argument for this approach to normativity, it’s at least possible (I hope you’ll agree) that pursuing theoretical questions through this practice-focused lens will turn out to be illuminating. So, short of offering a treatise on methodology, which I have neither the space nor—frankly—the wherewithal to do, I’m going to try pursuing certain questions about reasons from this currently somewhat unusual starting place. It is not as if other starting places have proved frustration-free, after all, or have satisfied all comers.¹³

¹² Or indeed the myth of atoms as indivisible, or there being just one kind of sub-atomic particle which in turn comprise atoms.

¹³ Maybe I should just say: they haven’t satisfied me. For one thing, I’m not convinced that we gain much of an explanatory advantage by starting our story about practical normativity with the concept of a reason—insofar as that is indeed the hope, which may be a partly question-begging assumption. In any case, the worry here would be that our grip on the Parfit–Scanlon concept of a reason may be no more or less secure than our grip on the concept of practical normativity itself, the former being essentially the stipulated minimal unit of the latter.

In casting around for a way to begin our story about practical normativity, starting from an account of human practices and activities, one could do worse than to consider P.F. Strawson's influential essay, "Freedom and Resentment" (1962/2008). There, Strawson drew an important distinction between the interpersonal and the objective stances which we can take to other people, from the second-person point of view. It's perhaps easiest to understand this distinction—at least in broad outline—by focusing initially on what the interpersonal mode is. When we interact with another person in this mode, we view them and treat them as a human being much like ourselves. (As a fellow human being, would be another way of putting it.) For one thing, we regard them as a sovereign creature—or, roughly and less fancily, as both equipped and entitled to make their own decisions. Moreover, absent further facts about our legal or social relationship, we generally regard ourselves as having no particular practical (as opposed to epistemic) authority over them. We are presumptively moral and social equals, capable of having a 'civilized conversation,' provided there are no contingent practical barriers, such as speaking different languages. And we are mutually intelligible, and regard ourselves as such.

This is a quick and dirty sketch of what the interpersonal mode of interaction paradigmatically consists in. (And note that I make no claims to perfect exegetical accuracy; I am happy to embroider Strawson's ideas in my own skein.) The bases on which we retreat from the interpersonal mode are many and various, however. It would be inappropriate—or simply impossible, Strawson sometimes seems to suggest—to remain in the interpersonal mode when we recognize that we are dealing with someone who is 'out of their mind,' or someone who is severely intoxicated.¹⁴ In both of these sorts of cases, we view the individual as a kind of human object to be managed, cured, or navigated around. And we may adopt the objective stance when we are dealing with young children as well. Here, we view our charges as in need of patience and also training—or, perhaps better, an education. Strawson also mentions the idea of taking the objective stance to someone as a 'retreat' from the emotional burdens and vicissitudes of genuinely interpersonal interaction. He made it clear that he thinks we tend to do this pretty regularly.¹⁵

¹⁴ I hesitate to use the expression 'out of their mind,' but there are few completely inoffensive ways of getting quickly at what Strawson is envisaging here. ("An idiot, or a moral idiot," is how he himself puts it; 1962/2008, p. 13) It may also be inappropriate to adopt the interpersonal stance when we are dealing with people who have certain serious intellectual or emotional incapacities. It might be responded that the interpersonal mode is more flexible than that, and can be adjusted to different levels of cognitive and emotional well-functioning. For my purposes here, I can afford to leave these potentially delicate and politically contentious issues open.

¹⁵ Although one wonders if taking the objective stance to someone who we could take the interpersonal stance towards is part of what some Kantians think that we should never do. And one wonders just how difficult that might turn out to be. Moreover, there will clearly be different (and more and less humane) ways of viewing and treating people in the objective mode. What should we say about how, as well as when, to take the objective stance, then? How are we to take the objective stance towards someone without objectifying her unduly? And is the objective mode just one mode or rather many? It is not as if we view inebriated adults as being children—let alone vice versa. These are important questions to discuss on another day though.

How exactly to think about the interpersonal and the objective stances—as well as how to draw suitable boundaries around them—are large and difficult questions, which I can't do justice to in this context.¹⁶ But fortunately, for my current purposes, it will be enough to look at how various emotional and relational possibilities belong on different sides of the divide. Strawson famously contended that it is only in the interpersonal mode that it is appropriate (or perhaps, even so much as possible) to adopt the reactive attitudes towards someone—such as resentment and also gratitude. He also held that the reciprocal love which two adults may feel for each other is only possible in the context of relationships that unfold (largely?) in the interpersonal mode. What I need you to accept, going forward, is the following of Strawson's suggestions: it is only insofar as one adopts the interpersonal stance towards someone that it is possible to reason or argue *with* them, or offer them collaborative advice or friendly suggestions, about what they ought to do.¹⁷ As Strawson elegantly puts it: “If your attitude towards someone is wholly objective, then though you may light him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.” (1962/2008, p. 10)

Why think as much? My first reason for thinking as much consists in a blatant appeal to authority: Strawson himself said so. My second reason for thinking as much consists in an appeal to your intuition. Imagine trying to have an argument with someone about what he is doing, and realizing that he is not someone to whom adopting the interpersonal stance currently makes sense. For example, he is out of his mind, or is blind drunk, or is actually aged three. In such cases, it seems to me that there is a meaningful sense in which we do and should switch into a different ‘mode’ of interaction.¹⁸ For example, we start working out how to extricate ourselves from the situation, with or without getting him to do what we originally wanted him to do. It is natural to talk about this relational shift in much the way that Strawson suggests. That is, we give up on the idea of arguing it out, or reasoning with him, and start to do something else. Perhaps we start to talk (bark?) *at* him,

¹⁶ Moreover, I will tend to speak for simplicity as if the interpersonal and the objective modes are mutually exclusive and exhaustive, with regards to particular issues on which we try to engage with people, morally. This is doubtless too simple, as Strawson clearly recognized. Most of human life is lived in the complicated middle, and also in epistemically uncertain relational territory, in which we have to ‘feel our way.’ I'll flag various other complexities that crop up as we continue.

¹⁷ I am using all of these various expressions to gesture as best I can towards the social practice I have in mind here. But there is no perfect term for it, so feel free to pick and substitute your favorite. And you should also feel free to hear these expressions slightly differently, or use them to mark distinctions internal to this domain. But their fairly close relationship is evinced by the possibility of very similar sorts of parodies—as in the menacing turns of phrase: “Let me give you a friendly piece of advice” and (says the gangster) “We had a little conversation” or (intoned ominously) “I can be quite persuasive.”

¹⁸ Again, I am trying to remain neutral on the question of whether this switch is a normative or conceptual mandate, a point on which Strawson (as I read him) is not entirely clear. But it makes little difference to my argument, I believe, whether we cannot take the interpersonal stance to those who we recognize as being indisposed in some way, or whether it is merely that we would be making a mistake in doing so. Either way, the relevant practices, roles, and stances plausibly involve a form of normativity which is internal to them, and is something like the normativity of rule-following *writ large*. Thus, the resulting practice-based approach to normativity might well be naturalistic but in a certain sense non-reductive. I explore these issues in work in progress.

where previously we were trying to speak to (or, better, with) him. Alternatively, we start to wheedle and cajole him, where previously we were trying to appeal to his ‘better nature,’ his capacity to be reasonable. We may maneuver him in ways that are more or less literal. Or we simply walk away, and give up on the intended intervention.

I suggest that the following claim has emerged as intuitively plausible at this juncture: reasoning with someone belongs to the interpersonal mode of interaction.

I now want to make the following, ostensibly modest proposal. We can naturally think of the reasons for an agent to perform some action as those considerations which would ideally be apt to be cited in favor of that action, when we are reasoning with her about what she ought to do.¹⁹ That is, these considerations would *actually* be apt to be cited in such contexts in which the agent is interacting with her (fictitious) ideal advisor, who is performing in this role as well as is humanly speaking possible.²⁰ This advisor should be imagined to be a flesh and blood human being—as opposed to the disembodied voice of reason—who is ideally suited to play this social role, partly in being possessed of all the relevant information and fully procedurally rational (or at least as fully procedurally rational as any actual human being could be). She might also be imagined to be virtuous and wise, perhaps—and she must at least be well-disposed towards her advisee. Finally, she should be imagined to be especially well-suited to play this social role for the particular agent in question. She is the person who is best suited to ‘getting through’ to her, morally.²¹ For, most people are much more amenable to taking advice into consideration when it comes from certain quarters—although it is important that they take it into consideration *as* advice, rather than implementing it merely in order to please this enigmatic advisor figure. For, as I’ll go on to explain in the next section, this would not be uptake of the sort that is aimed at by the social practice which I am envisaging. The role of the ideal advisor is to persuade or to recommend, not to issue *de facto* commands to the deliberating agent.

¹⁹ It is important to remember throughout that we are talking about reasons for action here (see *n.* 1). For, on a practice-based approach, the nature of reasons for belief and desire (in particular) will be very much an open question, whose answers will depend upon the contours of the relevant critical practices. I want to remain neutral on these issues in this context, but thanks to Tyler Doggett and Miriam Schoenfeld for helping me to think about them further.

²⁰ The figure of the ideal advisor plays an important role in Peter Railton’s discussion of the agent’s own good (1986), Michael Smith’s ‘advice model’ of internalism (1994), and also picks up on various remarks of Williams’ (of which more shortly). But I take it that my approach is more explicitly social in its emphasis than approaches like Railton’s and Smith’s, in particular, which bill the ideal advisor as being an idealized version of the deliberating agent herself. Although I think it is possible that some people are their own best advisors (if they are particularly resistant to taking an outsider’s advice on board), I see no general reason to restrict things in this way. And Smith seems to be thinking of the advisor as merely a metaphorical device, whereas I want to be considerably more literal-minded about it.

²¹ There are various complications here which I am setting aside for the sake of simplicity, since I do not need to settle them for the purposes of the discussion. For example, some agents might be more responsive to advisors who are or at least represent themselves as being fellow sinners. And whether or not the advisor must be entirely truthful with the advisee seems to me a substantive normative question about the standards of best practice here. Moreover, as well as the intended implication that the ideal advisors for A and B might be very different people, there is also the possibility that different people can get through to A maximally well regarding different particular matters.

The above proposal about what reasons are seems to me attractive partly insofar as it secures a close connection between reasons for action and the activity of reasoning with a person about what she ought to do.²² It is hard to believe that the entities and the activity could come too far apart. Surely the connection goes deeper than the common etymological root of the corresponding English words.

Another advantage of this proposal about the sort of job description which we should reserve for the notion of a practical reason is that it comes fairly close to Bernard Williams' own usage. And, since my ultimate aim is to defend a nearby relative of Williams' internalist position, there are obvious advantages in terms of clarity in sticking close to Williams here. For, although it is not always recognized, Williams was explicit about the fact that he saw practical reasons as a *distinctive* normative notion. He wrote, for example, that "From both an ethical and a psychological point of view it is important that 'A has a reason to X' and its relatives should say something special about A, and not merely invoke in connection with him some general normative judgment."²³ (1995b, p. 192) Elsewhere, we find clues about the nature of the intended contrast here. Take Williams' discussion of a man who broke an obligation or violated someone's rights. Williams wrote that "perhaps he had no reason at all [not to do so]. In breaking the obligation, he was not necessarily behaving irrationally or unreasonably, but *badly*. We cannot take for granted that he had a reason to behave well, as opposed to our having various reasons for wishing that he would behave well."²⁴ (1985, p. 192, my italics) The idea seems to be that there is simply bad behavior, on the one hand—such as might be exhibited by a madman, a drunkard, or a child, perhaps. That is, there is behavior of the kind which we ourselves have reasons to discourage them from engaging in, since it has bad effects on others (say) or is otherwise anti-social. But there is also the distinctive category of unreasonable behavior, on the other hand—which an agent as such has reasons as such to refrain from.²⁵

²² One can undertake this activity alone, of course, but I'm inclined to think of individual deliberation as a kind of conversation with yourself, in which you are playing the dual role of advisor and advisee. Thus, the individual activity is in my view parasitic on the relevant interpersonal practice. For a discussion of the possibility of taking the analogue of the interpersonal and the objective stances towards oneself—as Strawson assumes is possible—see my (2014). Thanks to Alex Guerrero for prompting me to be clearer about this.

²³ And he went on to argue that "internalism in some form is the only view that plausibly represents a statement about A's reasons as a distinctive kind of statement about, distinctively, A." (1995b, p. 194)

²⁴ But lest it be suspected that obligations or rights for Williams are supposed to be matters of mere convention, it is worth noting that Williams spoke in the same breath of some rights violations as 'monstrous.' As in: "...some of the most monstrous proceedings, which lie beyond ordinary blame, involve violations of basic human rights." (1985, p. 192) Williams goes on to oppose blame (which "seems to have something special to do with the idea that the agent had a reason to act otherwise") to other "ethically negative or hostile reactions to people's doings (it [being] vital to remember how many [of these] there are)." (1985, p. 193)

²⁵ Compare Williams' telling remark that, when reasons are not relativized to the deliberating agent's motivations, these 'external' reasons-claims: "...mean something that could be expressed by a different kind of sentence, for instance to the effect that it is *desirable* that A should do the thing in question, or that we have reason to desire that A should do it. Only the internal interpretation represents the statement as distinctively a statement about A's *reasons*. Relatedly, if a statement of this kind is true, and A declines to

Williams is also plausibly read as thinking of reasons as connected with our practices of reasoning with each other, as well as by ourselves. For, he spoke of questions about reasons as “uncontentiously connected” with the advice that we should give to an agent in the “‘if I were you’ mode” of assisted deliberation (1995a, p. 36). Similarly, he distinguished reasons as bearing on the question of what one ought to do (all things considered) in the distinctively ‘deliberative’ or ‘practical’ sense of the term ‘ought’—which he at one time contrasted with a moral and also a ‘general propositional’ ‘ought,’ as in “It ought to be the case that A does X.” (1981b, p. 118) Shortly afterwards, he characterized the practical ‘ought’ as pertaining to “contexts of advice or of discussion about what it is reasonable for an agent to do.”²⁶ (1981c, p. 125) Now, I don’t want to insist that the practice of collaborative advice-giving that Williams had in mind here is exactly the same thing as the mode of interaction which I’ve called reasoning with someone.²⁷ And it’s not as if I’ve offered a particularly sharp characterization of that activity (I’ll say a bit more in due course).²⁸ Still, there are obviously resemblances, and this is all to the good, in terms of getting clearer on the issues central to this paper.

But while the above proposal about how to understand the notion of a reason for action seems to me to be natural, and has the recorded conceptual and dialectical advantages, I do not want to pretend that it is at all uncontroversial. (A modest proposal indeed, then.) On the contrary, it involves rejecting the popular suggestion that practical reasons are just the basic unit of practical normativity, our equivalent of a (or indeed, rather, *the*) sub-atomic particle. For, whenever someone is doing something which is in any way bad or criticizable, those who envisage reasons on the sub-atomic particle model will tend to want to say that this person has a reason to cease and desist.²⁹ Whereas I have suggested that we construe reasons

Footnote 25 continued

do the thing in question, what is called into question is A’s capacity in this connection to act *rationally* or *reasonably*.” (1996, p. 109, my emphasis) But Williams later conceded to Scanlon that his opponent need not hold that someone who flouts a supposedly valid external reasons-claim should be described as being irrational per se, as opposed to merely unreasonable (2001, p. 93).

²⁶ He also goes on to clarify that “...an adviser may say that A ought to do X and, at least if the adviser speaks in the mode of relative practical advice, he surely says the same thing as A would say if A said ‘I ought to do X,’ and something that would be contrary to A’s saying ‘I ought not to do X.’” (1981c, p. 128) Later on, he reiterated that “The stance towards the agent that is implied by the internalist account can be usefully compared to that of an imaginative and informed advisor, who takes seriously the formula ‘If I were you...’” (2001, p. 94) See also 1981b, p. 120, 1985, p. 193, and 1995a, pp. 40–42. But note the unfortunate exception of the original 1981a piece here, save for a passing reference therein to the “persuasions of others.” (1981a, p. 105)

²⁷ Although Williams did distinguish, in a broadly Strawsonian vein, “between two possibilities in people’s relations. One is that of shared deliberative practices, where to a considerable extent people have the same dispositions and are helping each other to arrive at practical conclusions. The other is that in which one group applies force or threats to constrain another.” (1985, p. 193) Congenially, he implies that reasons-talk belongs within the first of these relational possibilities.

²⁸ Moreover, while I am inclined to think of the individual activity as parasitic on the social one (see n. 22), Williams does not seem clearly committed one way or the other.

²⁹ Parfit: “We cannot criticize or blame people for failing to do what we believe that they have no reason to do.” (2011, vol. 2, p. 442) I believe that Parfit is wrong about the criticism part, although he may be right about the blame part, as I’ll suggest in Sect. 3.

significantly more narrowly. If one adopts my proposal, then we will say that someone has a reason to do otherwise only insofar as this person is behaving not only badly but unreasonably, in the sense that it still makes sense to adopt the interpersonal stance towards her and to try to talk her out of it (or to remonstrate with her after the fact in a reasonable tone of voice, as it were). That is, a reason for an agent A to ϕ is a consideration which A could be expected to be responsive to if she was being reasoned with as well as is (humanly) possible. No such thing will be true if we think of reasons as being the sub-atomic particle of practical normativity—i.e., as comprising the minimal and also generic chunk of matter(ing).³⁰

Still, it is natural to wonder whether this issue might turn out to be at least partly terminological. I think we need to be careful here. Suppose that someone thinks of reasons as being the basic contributory notion when it comes to any normative phenomenon, but is open to the idea that there may be fundamentally different *kinds* of normative reasons—apples and oranges, as it were. And she is persuaded by my argument (I say hopefully) to distinguish a special class of reasons—call them ‘interpersonal reasons’ or some such—and goes on to say all of the things about them which I say about reasons *simpliciter*. In particular, she endorses a form of internalism about interpersonal reasons. I agree that this might merely represent a different choice of terminology—albeit, in my view, a somewhat less perspicuous one. But whether or not this is so would depend upon whether this theorist would indeed be prepared to grant that interpersonal reasons are importantly distinctive from reasons of other kinds. This is the issue which I take to be the heart of the matter, and which I think is clearly substantive.³¹ Namely, does the distinction between unreasonable and simply bad behavior (whether or not we use the word ‘reason’ to encompass behavioral criticism of both kinds) really cut very deep? Theorists who subscribe to the sub-atomic particle model would presumably want to deny that there is any deep distinction here, when it comes to the normative claims that hold true with regard to an agent. Whereas practice-based theorists should at least be open to affirming the depth of this distinction. For, the practice of reasoning with a person about what she ought to do might turn out to involve considerably more than just identifying what it would be good or desirable for her to do, from a social point of view (say).³² In the next section, I will be arguing that this is in fact

³⁰ Parfit and Scanlon would want to say, admittedly, that a reason for an agent A to ϕ is a consideration which A could be expected to be responsive to if she was fully *substantively rational*. But, as will become clearer in the next section, this is a far weaker constraint on an agent’s reasons than the condition I intend here—if it should even be understood as being a constraint at all. It should plausibly be read as going the other way, i.e., as representing a constraint on one’s account of substantive rationality. Thanks to Reid Blackman and Daniel Star for discussion on this point.

³¹ Thanks to Julia Markovits, Sarah Stroud, and Ted Sider for pressing me to be clearer about my own commitments here.

³² Although it is plausible to think that it will be highly sensitive to such issues. One might think, for example, that reasoning with a person aims to get her to do something which is at least acceptable or ‘good enough.’ However, a crucial complication is that, for very bad agents, we as their advisors should arguably try to persuade them to do things which are still quite bad, but less so. And yet it does not seem right to say that we should try to get the agent to do things that are the best of a bad bunch of options

the case. In particular, I'll be arguing that reasoning with a person requires a certain sort of *receptivity* on her part to the relevant considerations, which plausibly entails a certain sort of disposition to be motivated to act on such a basis. Whereas we can surely evaluate an action negatively without presupposing such responsiveness on the part of the agent who performs it. What we might be saying is that the action is unacceptable according to standards we endorse, or that it would be bad from the perspective of members of the moral community who we are presuming to speak *for*. Or we might be issuing the agent with a kind of moral report card about the quality of her behavior. We would not necessarily be attempting to bring her in and speak with her about what she has been doing—unless we assume that, like many of us, she cares about earning a decent grade. I'll come back to such possibilities later in the paper.

Thus, from a practice-based perspective, there is essentially a risk that the notion of a practical reason that Parfit and Scanlon are working with will represent a gerrymandered category, or lump together apples and oranges. For, this notion may elide an important distinction between various critical practices which should rather be brought to the fore. Parfit and Scanlon say that a reason for an agent A to ϕ is a fact which counts in favor of A's ϕ -ing. The question you should hear me as pressing is this: counts in favor *how*, i.e., in the context of what sorts of human relationships? Part of what is at issue here is that I do not really think that facts count in favor of doing anything whatsoever. Indeed, in a certain way, I do not see how they could. Surely 'counting in favor' is a distinctively human activity—something that people do by citing facts, more or less appropriately, given the relevant social context or the sort of game we're playing.³³

It may be helpful to pause on the back of this to note what I'm not up to in this paper. Theorists like Parfit and Scanlon are often taken to task for being non-naturalist realists about reasons. In one popular unflattering metaphor for their picture, reasons are somehow "out there," and this is supposed to be incredible. But reasons could be out there in a number of different senses. For one thing, they could be out there metaphysically, in the sense of being irreducibly normative entities (or grounded in irreducibly normative truths and facts in a more complicated way). Or, reasons could be out there in the sense that they are pictured as explanatorily prior to human practices and activities, with human beings depicted as trying (or, perhaps, failing) to orient themselves to their dictates, which regulate and could theoretically float free of human social life. These two pictures often go together, but I think their relationship to each other (and with naturalism) is quite complicated, and may to some extent be disentangled. In any case, it is the second picture rather the first which I am questioning just now. I think that human beings and their activities

Footnote 32 continued

which she might be persuaded to take. If the best of these options is not only bad but *terrible*, the ideal advisor might be called upon to simply walk away from the whole sordid business. I do not have to take a stand on these issues in this paper. But suffice it to say, I think that the relationship between the advisability and the desirability of some arbitrary action will tend to be quite complicated. And I think of the agent's good character as the glue that will often serve to hold the two together.

³³ Compare J. David Velleman (ms) for a similar sort of view.

should come first in the order of explanation, when it comes to practical normativity and morality more especially. For an oversimplifying slogan: reasons do not make claims on us; rather, people do.³⁴ So that is the perspective on normativity which I'm attracted to exploring and trying to bring into view here. I think that internalism about reasons is one of the many upshots of looking at things from this angle. Or so I'm about to argue.

2 Motivations and the limits of the interpersonal stance

In the previous section of this paper, I mentioned some uncontroversial examples of cases in which we are called upon to shift into the objective mode: when we learn that the individual is (in some sense) not in 'their right minds,' enduringly or temporarily, or when it turns out that the individual is still just a child. In this section, I'm going to argue for a more contentious claim about the retreat from the interpersonal mode. Namely, I think that we fall out of the interpersonal mode (and thus, out of reasoning with someone) when we ignore certain facets of our interlocutor's motivational profile, and try to get her to do things which she is not at all motivated to do, and would not become motivated to do, simply as the result of our continuing to talk and reason with her. Coupled with the suggestions of the previous section, this claim will evidently push us in the direction of reasons internalism.³⁵

Let me bolster my target claim by again appealing initially to certain of your intuitions. Suppose, once more, that I am trying to get someone to alter his behavior. For example, to borrow Williams' well-known example, I am trying to convince a man who is nasty to his wife to treat her more nicely, or with more consideration. In this endeavor, I repeatedly press my concerns on him, and in a variety of ways. Finally, he says to me—borrowing Williams' wording here—"I don't care. Don't you understand? I really do not care." (1995a, p. 39) That is, he doesn't care directly about treating his wife more nicely. Nor does he care about any of the goods which would be promoted or instantiated by so doing. He doesn't care about improving his wife's well-being, his marriage, or even his own lot on this score (assuming, as is plausible, that the well-being of the two partners is not wholly unconnected). He is not interested in being a good husband in the abstract. Nor does he care about acting

³⁴ *Pace* Parfit, who writes that, when we ignore the facts which give us reasons, "we are not responding to them, just as ignoring someone's cry for help is not responding to this cry." (2011, vol. 1, p. 32)

³⁵ In offering a Strawson-inspired argument for reasons internalism, I am admittedly distancing myself from Williams' official argument for the view, which has to do with possible explanations of an agent's actions. (See Williams 1981a, in particular.) Like many others, I find this argument quite opaque, and am not convinced that it captures the best way of motivating the position. There is plausibly some connection to be found, which will crop up when I talk about the activity of reasoning with someone as an attempt to get her to act out of her recognition of the reasons we cite in favor of a certain action—and by means of which her subsequent action might be thus explained. But this connection is not straightforward, owing partly to the possibility of considerations aptly adduced in advice which the agent might be duly motivated to act upon but which would never actually result in action, due to this motivation being accompanied by a contrary and overriding motivation whenever it occurs.

more kindly and considerately in general. He is not even moved by the thought that his wife is a fellow human being, who will be hurt by the shabby way in which he continues to treat her. And so on and so forth. We are not getting anywhere.³⁶

Here's the intuition I have, and want to invite you to share, now: when we learn that this man cannot be motivated to lift his game merely by continuing to carry on with the conversation, something has now changed in the normative and dialogical space between us. (This is admittedly sketchy, but we'll get more precise as we go.) Entering into this exchange, I had hoped to have what we sometimes call a 'rational conversation.' I had hoped to speak to this man as one reasonable person to another, on behalf of his unfortunate wife.³⁷ I had hoped to get him—again, using a suggestive if vague expression we sometimes reach for in these contexts—to 'see reason,' vis-à-vis his conduct. I had hoped that he would, firstly, be willing to listen to what I had to say, and, secondly, that he would be willing to try to mend his ways.

I had hoped for all of these things in vain though. At this point in the exchange, if this man is to be believed—as I propose we assume for the purposes of the discussion—then there is nothing in his motivational cache (so to speak) which I can appeal to, to spur him into action. And, again, there is an intuition that this fundamentally changes our sense of what we are doing here, morally and socially. It seems to me that we can no longer correctly claim to be reasoning with this man about his treatment of his wife. Insofar as we keep hoping to alter his behavior, rather than simply backing off, we are no longer playing the interpersonal game of giving and asking for practical reasons. We are embroiled in an importantly different kind of interaction (of which more shortly).

But first, we need to get clearer on what prevents us from reasoning with this man, more precisely. Suppose there was a non-normative fact of some kind which we could point out to him, which would change his mind on this matter. For example, if we pointed out that his wife has been sinking into a depression on account of his behavior, or that she has been lingering at work just in order to avoid him, then he would be moved to treat her with more consideration. In this case, it seems intuitively plausible to say that we can continue to reason with this man about his behavior, even though it takes an injection of non-normative information to produce a motivational response in him which

³⁶ Remember that we are supposed to be imagining ourselves playing the part of the ideal advisor here, who may be presumed to know where this is going (i.e., nowhere). But an important complication is whether and how we can determine in practice that the person we are dealing with really can't be reasoned with. Perhaps they are absorbing more than they are letting on, or perhaps they will remember our advice and be receptive to it later. (As Alex Guerrero rightly pointed out in his BSPC commentary, "The arc of the moral universe is long.") I agree that it's hard to know in practice whether the person who we're dealing with is genuinely non-responsive. I also believe that we are often obliged to assume—on analogy with the principle of charitable interpretation—that our interlocutor is capable of being reasoned with until we have something approaching knowledge that this is not the case. But we can imagine in cases like the above our having reached such a point, at least if we were to know the callous husband 'inside out.' Thanks to Larisa Svirsky and Alex Guerrero for helping me to think through these issues.

³⁷ It is natural (disturbingly natural) to describe and imagine this interaction as a conversation "man to man"—especially since the perspective of the wife never enters into it in Williams' original discussion, and is similarly elided in much of the ensuing literature. That she does not seem to have a voice here is well worth reflecting upon, on a number of different levels. I'll leave these reflections for another day though.

is currently quite lacking. Such examples make it tempting to say that we should consider not this man's actual motivational profile, but the motivational profile of a well-informed and (it is also often added) procedurally rational counterpart of the person.

I think we should be cautious about this popular line of thought, though. For one thing, it seems that not just any injection of non-normative information will do: if what it takes for this man to lift his game is the *non sequitur* (albeit true) that his wife's secretary was wearing a yellow tie today, then we would hesitate to say that the process of reasoning with him is continuing unabated. Rather, it seems that there is some weird feature of this man's psychology which is functioning arationally, much like the proverbial blow to the head. And suppose now that we substitute a normatively pertinent claim—such as the fact that his wife has been spiraling into a depression on account of his perpetual nastiness. Still, we have to specify that the role this belief plays in inducing a change to this man's motivations is not merely causal but rather somehow rational. That is, his motivations change in view of his sense of the normative significance of said fact.

So we might now consider saying, borrowing a leaf from Williams' book, that reasoning with someone is constrained by the motivations which he might reach, on the basis of his engaging in sound practical deliberation.³⁸ This process would involve the person becoming knowledgeable about the normatively relevant non-normative facts, and relieved of procedural errors in his reasoning, among other things. We thus consider a counterpart of the agent whose motivational changes have been induced by rational means.

But even this proposal is problematic, by the lights of the current perspective. Suppose, for example, that this man is unmotivated—and indeed, completely unwilling—to engage in sound practical deliberation about the issue at hand. It strikes me as something of a cheat to say that we can still reason with him about treating his wife more nicely, because he could be persuaded to do so if he deliberated soundly, if he could not be persuaded to consider the matter in the first place. Reasons internalists are often animated by the thought that we should avoid depicting people as essentially *alienated* from their reasons.³⁹ And the idea that reasons are the kinds of considerations with which we can reason with a person is a recognizable variant on this theme. But if it is not the person herself but rather her better counterpart who would be motivated to act in accordance with some supposed reason (be it following sound deliberation or no), then surely she is alienated much as she was before.

A suggestion, then: we might think of the process of reasoning with someone as constrained not by her actual motivations, nor by the motivations some idealized version of her would have. Rather, we could think of the process as constrained by the motivations which she *herself* would form, following a complete series of moves within an idealized version of this practice. Then, we—acting as what Williams called her “imaginative and informed advisor” (2001, p. 94)—would be constrained

³⁸ David Sobel makes a similar move at this point in a similar dialectic, in his helpful paper on reasons internalism (2001, p. 223).

³⁹ I believe the metaphor of alienation in the context of broadly internalist views is originally due to Railton (1986).

by motivations which the agent would have at the end of the conversation, once we had finished sharing pertinent non-normative information which she is actually willing to hear about, and offering those corrections to her reasoning which she is receptive to receiving.⁴⁰ We might do other things as well here, such as kindling her imagination, Williams thought (more on this in a moment). More abstractly, the process of reasoning with someone will be depicted as constrained by those motivations which the agent herself would have if the process of reasoning with her were to be perfected and completed. We are idealizing not the agent, but the relevant interpersonal process.

Many more details could be adduced here, and a full defense of my views would certainly require them.⁴¹ But you can see their basic shape. In particular, you can now see how the version of reasons internalism I'm developing will be similar to Williams' in some respects, but at least potentially different in others. I argued in the previous section that reasons for an agent A to ϕ are considerations which would be apt to be cited in favor of A's ϕ -ing, when we are reasoning with her about what she ought to do, as her ideal advisor who is doing her job as well as is possible. If what I have suggested in this section so far is along roughly the right lines, then a consideration would only be apt to be cited in favor of A's ϕ -ing in this context if A would be (at least somewhat) motivated to ϕ , by the end of the conversation. It follows that the reason for an agent A to ϕ can hold only if A would be (again, somewhat) motivated to ϕ , at the end of an idealized process of being reasoned with in this way.⁴² She must have the relevant motivational *propensity*, is a convenient

⁴⁰ Following Williams, I see no pressing need to deny the possibility of a certain amount of indeterminacy about what the outcome of such a conversation would be, which would subsequently be inherited by the notion of a reason (1981a, p. 110).

⁴¹ In particular, we need to be very careful about how we individuate conversations, in order to prevent some of the well-known conditional fallacy worries which afflict 'ideal agent' models from afflicting this account too. For, we need to simultaneously do justice to Williams' thought that a man about to drink a glass of petrol doesn't have a reason to do so—because he could be quite easily talked out of it, simply by pointing out that it is not quite the gin and tonic he hoped for—while also doing justice to the thought that the man has a prior reason to inquire into the contents of the glass. The second reason is tricky because it is so transient (or, as Sobel calls it, 'fragile;' 2001); the better informed man clearly lacks it. I would be inclined to try to deal with these issues by proposing that conversations should be individuated in a more fine-grained way. We would then say that the ignorant man has a reason to acquire more information, on the proviso that he would be interested in doing so. ("Do you want to know what's really in that glass?" we might ask. "Yes," he would then reply.) But, time is pressing, and we very reasonably presume that he does have this reason. So we tend to cut out the middle man, and simply offer him the information (by saying "That's petrol!"). The informed man has a reason to refrain from drinking the contents of the glass, insofar as he has been successfully persuaded on this point and is hence motivated to refrain. And he always had such a reason, because there was always a conversational path which he could be persuaded to take which would persuade him not to drink up. Whereas the reason to acquire more information is duly transient, because the willingness to acquire it only survives the first 'sub-conversation' in the overall exchange just described. Of course, this is only the briefest sketch of how one might deal with a delicate and important issue.

⁴² Things evidently become more complicated when we consider that reasons come in different strengths. And we might think that the advisor's voice should get softer in response to a proportionally weak motivation, as well as falling completely silent in its absence. Thus, I think it is natural to say that, just as a person's ultimate lack of motivation to do as we recommend defeats the (*pro tanto*) recommendation which we would otherwise have made to her, so a proportionally weak motivation to ϕ diminishes the

way of putting it. This seems likely to lead to an even stronger necessity condition on reasons than the one that Williams defends. For it seems plausible, on the face of it, that there might be ‘sound deliberative routes’ which we as an ideal advisor could not talk some agents into taking.⁴³

But there are also important similarities between Williams’ view and the version of reasons internalism that emerges from this argument. For one thing, it does not generalize very rapidly from a necessity claim about reasons to a sufficiency claim as well—which echoes Williams’ insistence that internalism about reasons should be understood as being a necessity claim only (see *n.* 8). However, as we saw there, Williams did say that he found the sufficiency claim to be plausible. Whereas I am inclined to think that there is nothing especially plausible about the corresponding sufficiency claim—i.e., that there is a reason for an agent *A* to ϕ if *A* would be motivated to ϕ , at the end of an idealized process of being reasoned with in the relevant way. For, it seems to me implausible that we as her ideal advisor should recommend her doing whatever she would be motivated to do at the end of the conversation. Perhaps some such claim will hold vis-à-vis the agent’s own well-being, or what she should do for her own sake. But I doubt that such a claim is true in general, and it would certainly need to be argued for. So subjectivism is far from inevitable, if one has my motivations for espousing reasons internalism. The position may remain a necessity claim only.⁴⁴

The version of reasons internalism I’m endorsing is thus unusually strong in one way and unusually weak in another (or somewhat unusually, anyway). And it at least has the scope to be unusually weak in another respect too. For nothing I’ve said

Footnote 42 continued

strength of the relevant recommendation, and hence the corresponding reason, *pro rata*. But I will not try to settle this issue here. Thanks to Geoff Sayre-McCord for helping me in thinking about it.

⁴³ I don’t think Williams would have agreed with this. He is explicit about his assumption that agents are virtually always interested in getting straight on the facts (1995a, p. 37), which seems to me too optimistic. This helps explain why he slid back and forth between talking about reasons as constrained by motivations which the agent himself could reach (see, e.g., 1995a, p. 35), and talking about reasons as constrained by motivations to which there merely exists a path, via a sound deliberative route from the agent’s existing motivations (see, e.g., 1995a, p. 36; 2001, p. 91). However, Williams did once suggest that, when it comes to understanding the somewhat opaque notion of a ‘sound deliberative route,’ we might “reverse the order of explanation, and, in some part, place the constraints on the procedures that are to count as deliberative assistance in contrast to these other interventions [such as manipulation]. What someone has reason to do will be what he can arrive at by a sound deliberative route; and he can arrive at a conclusion or resolution by a sound deliberative route, perhaps, only if he could be led to it by deliberative assistance that operated within those constraints.” (1996, p. 115) This is very close to the kind of idealizing I’m proposing we go in for here. I would just add that this idealized process may require not foisting information on the agent, or otherwise questioning her judgment, if she simply doesn’t want to hear it.

⁴⁴ Relatedly, I think that my version of reasons internalism lends little credence to the idea that an agent’s motivations are the *source* of all of her reasons, in either of the two ways that the ‘source’ metaphor is commonly understood. The sorts of considerations that provide reasons (or potential reasons, subject to being enabled) could be anything whatsoever, even if these reasons would effectively be defeated or ‘vetoed’ by the agent’s lacking the relevant motivational propensity. Moreover, when it comes to explaining *why* these sorts of considerations are fit to provide reasons—i.e., why they at least have the potential to have genuine normative force—part of the explanation here would presumably be that they are potentially apt to be cited in the relevant critical practice.

so far rules out the possibility that the process of reasoning with someone must be beholden to what is there in the agent already, by way of motivation. Consider the thought that we as an advisor can aim to inspire or challenge the agent in ways which would produce in her genuinely novel motivations which she did not have before. We might do this by persuading her to—just as Williams repeatedly mentioned—use her imagination, engage in new sorts of experiences, or—as Williams notably omitted to mention—engage with *people* with different backgrounds, proclivities, and perspectives. We might also encourage her to read a certain novel, or get ‘out of her own head’ in any number of different ways. What seems important to me is just that we are leading her down a path which she is interested in going down. We are thus picking up on motivational propensities properly ascribable to her, rather than essentially trying to mold her into a different kind of person.⁴⁵ So my sense is that inspiration may have a place within the interpersonal activity of reasoning with an agent. There at least seems little basis for ruling this out at the outset. But even if we ended up relaxing the internalist constraint on reasons in this way, it would still be a highly non-trivial constraint, having been tightened in another: namely, via the insistence that the agent must be open to the information and experiences needed to germinate the seeds of these novel motivations. There are many practical matters in which we are not so open—happily or unhappily. For example, I am quite confident that I couldn’t be persuaded to convert to scientology over the course of a brief discussion about what to do with my day. And if you were to know me and were told that this had happened, you could deduce that the so-called conversation must have devolved into something else, such as manipulation or brainwashing, rather than advice. It is not just that I am currently unmotivated to convert to scientology; it is that I am not even willing to hear out the sales pitch.⁴⁶

I’ve been suggesting as intuitive the claim that someone’s lack of motivational propensity to do as we recommend is a roadblock beyond which we cannot proceed when we are trying to reason with her about what she ought to do. Still, there is the question of why this might plausibly be the case. Here, I think we have to look closely at what the activity of reasoning with someone might be said to involve. Reasoning with someone is—I want to claim—an activity in which we are not only trying to get the agent to *do* something (or at least to contemplate doing it). Rather, I think we are trying to get the agent to do something (or, again, to contemplate doing it) out of her sense that this thing is actually worth doing (or somewhat worth doing, if the relevant reason is not decisive). Here’s how I’m tempted to develop this thought further: when we are reasoning with someone, we cite some consideration in favor of her ϕ -ing in the

⁴⁵ Compare Williams, who wrote—in a vague but arguably similar vein—that “we still need the notion of the decision being an expression... of motivations that the agent had in the first place... unless those motivations themselves are expressions of what was there before.” (1996, p. 116)

⁴⁶ Hence, my hunch is that, while inspiration may arguably have a place here, *bona fide* conversion is out—a conversation involving conversion being no conversation at all. This is *pace* John McDowell (1998), who argues that a consideration could be a reason insofar as the agent would be motivated by it, if he were to come to see the matter aright, having undergone a ‘conversion’ to reasoning correctly (rather than the transition having to be effected by correct reasoning from pre-existing motivations, which is how McDowell reads Williams). So I am friendly to what may in effect be an intermediate position here.

hopes that she will come to *recognize* it as a reason. This, and only this, would count as uptake of the aimed-for kind. It is not enough that she start φ -ing, in deference to us, say. We want the agent not just to *act out* a certain recommended action; we want her to *act out of* her recognition of what there might be to recommend it.

However, there is an important complication here which it is worth pausing to bring out.⁴⁷ It has to do with the fact that there is often more than one potential reason for an agent A to φ . Throughout this paper, I've been talking about reasons using the deliberate fudge word 'consideration,' so as to try to remain neutral on the question of what reasons are, exactly. But suppose for a moment—just for the sake of clarity—that reasons consist in facts, as it is commonly supposed. Then it might be that two facts both have the potential to be apt to be cited in favor of A's φ -ing, and A is thoroughly unresponsive to the one but not the other. Imagine, for example, an obtuse but basically caring husband. And imagine that he could not be prevailed upon to recognize the fact that his wife is mentally working through a tricky situation at work as a reason to listen to her attentively without jumping in immediately with various solutions to the problem. ("What is the point in that?" we can imagine him saying or thinking, in response to the advice that he pipe down and just listen—she needs to figure this out for herself.) But he does want to be a good spouse. And he is also prepared to believe, or simply to take it on trust, that listening in this way is part of what is involved in being a good spouse in these sorts of situations. It follows that the fact that good spouses sometimes just listen might provide him with a reason to do so, seeing as he could be motivated to recognize this fact as being a reason (or so we are imagining, anyway).⁴⁸ And this is the sort of uptake which I think that advice aims to achieve—wherein the agent recognizes some potential reason as being a reason, which may be held to cement its status as being a reason proper.

If we were to simply leave things there, then internalism about reasons would not follow. But there is a tempting—although certainly not uncontroversial—claim in the offing that, when someone fully recognizes that there is a reason to φ , then she will be motivated to φ , at least to a certain extent. The activity of reasoning with someone could then be said to have the aim of getting our interlocutor into a state wherein she will be motivated to do the things which we suggest she has reasons to do. This version of reasons internalism might thus draw strength from a version of motivational internalism by such means. Suppose you share the intuition that making some move in the game of reasoning with an agent will only be apt if this move would meet its aim if all goes maximally well—i.e., if the process of reasoning with her were to proceed in an ideal way. Then the implication (when

⁴⁷ Thanks to Julia Driver for raising it in her insightful commentary at BSPC.

⁴⁸ I say "might" though because, on my view, we would need to believe that the fact that good husbands do such-and-such is an independent, non-derivative reason to do the thing in question. I am sympathetic to this view in the case of social roles (see my 2013). But I am much less sympathetic to the view that there are independent, non-derivative reasons to be rational, or to act in accordance with the balance of one's reasons. So the *de dicto* desire to do so does not seem to me to threaten to trivialize reasons internalism (even supposing that such a desire is universal, which I would also move to deny). For, such a desire cannot enable considerations which are not eligible to provide (further, independent) reasons for action in the first place.

conjoined with my proposal about what reasons are) is that reasons hold for an agent only if she would wind up in a state in which she would recognize a reason and hence be motivated to ϕ , at the end of an idealized process of being reasoned with in this way. This being exactly the form of reasons internalism I endorse. Here is the argument laid out in proper detail:

(P1) A reason for an agent A to ϕ is a consideration which would be apt to be cited in favor of A's ϕ -ing, by her ideal advisor, who is reasoning with her in an ideal way about what she ought to do.

(P2) Reasoning with an agent about what she ought to do is an interpersonal activity which aims to achieve uptake—wherein the agent comes to recognize the cited consideration as a reason for her to ϕ .

(P3) In the context of reasoning with an agent about what she ought to do, citing some consideration as a reason for her to ϕ is therefore only apt if she would come to recognize this consideration as a reason for her to ϕ , if this activity were to proceed in an ideal way. (From P2)

(P4) Recognizing that one has a reason to ϕ entails being motivated to ϕ , at least to some extent. [A version of motivational internalism]

(C1) In the context of reasoning with an agent about what she ought to do, citing some consideration as a reason for her to ϕ is therefore only apt if she would end up in a state such that she would be (somewhat) motivated to ϕ , following an idealized process of being reasoned with in this way. (From P3 and P4)

Conclusion: An agent A has a reason to ϕ only if A would end up in a state such that she would be (somewhat) motivated to ϕ , following an idealized process of being reasoned with in this way. (From P1 and C1)

I think that this argument is interesting and worth pursuing further, although both the premises and the transition from (P2) to (P3) would require a lengthier defense than I can manage in this paper.⁴⁹ But, even if it went through, it would merely push the question elsewhere, in a way. Even supposing (P2) is true, *why* might it be true? Why is the process of reasoning with someone the kind of activity wherein we aim to achieve uptake and recognition of this kind?

This is a deep and difficult question, and I don't have anything like a full answer to it to offer. But here the idea of the interpersonal stance may again point us in the right direction, I think—in addition to providing conceptual support for the sort of democratic ideal which (P3) tries to capture. I suggested earlier that reasoning with someone can only take place in the interpersonal mode. Moreover, I suggested that the interpersonal mode involves seeing and treating someone as a sovereign creature, in the sense that she is both equipped and entitled to make her own decisions.⁵⁰ If we

⁴⁹ Although elsewhere, I defend a suitable version of motivational internalism such as would bolster (P4) from an (again) Strawsonian perspective (Manne 2014).

⁵⁰ Thanks to Japa Pallikkathayil and Michael Kessler for helping me to improve this point, which I had formerly expressed in terms of 'autonomy.' But it is a crucial part of my argument that we should group

connect the two suggestions, then we can offer the hints of the beginnings of an explanation as to why people's recognitional capacities and attendant motivational profiles (again, assuming a suitable version of motivational internalism) might matter so much here. Perhaps the issue is that we ignore someone's receptivity in this sense at the expense of no longer envisaging her as a sovereign creature, whose enduring unwillingness to do something is in a certain sense definitive. For, a reason belongs to the practice of calling each other to account. And a call requires the realistic psychological possibility of its being heard and answered, I think. Otherwise it does not address the agent as she is but rather says something more impersonal *about* what she is doing—or addresses an idealized counterpart of the person who is not actually on the scene. And we might also point out that we ignore someone's receptivity in this sense at the expense of no longer treating them as someone over whom we have no *prima facie* authority. Rather, we start to treat them as a moral inferior of sorts, insofar as we might be entitled to issue them with orders (or similar, or worse).

Which brings me to my next point. It may be entirely permissible or, indeed, obligatory to go ahead and issue people like the callous husband with orders (or similar, or worse). Doubtless this will depend on further details of the case. For example, it will depend fairly obviously on how bad his treatment of his wife is. (Is he merely callous, or positively abusive, is a good preliminary question.) There is also the pragmatic question of whether he is likely to comply with our orders. And we should carefully consider our relationship to him too. If he is a family member or a friend, we might be permitted to intervene much more strenuously in his life. If his wife is a family member or a friend, we might find it appropriate to intervene more strenuously still.

As Williams pointed out crisply, it is not always our place to make a moral "citizen's arrest." (1985, p. 214) And the question of how much we resemble the fictitious ideal advisor—particularly in terms of both our rectitude and our knowledge that we are indeed in the moral right here—will be very much an open one in many real life situations. Still, let me be clear here. I certainly do not believe that we have to retreat from doing or saying anything, if we discover that the callous husband has no motivational propensity to treat his wife more nicely. Rather, I believe that, if we discover that there is no way of motivating him to do so, simply by reasoning with him, then we are no longer well-described as trying to offer him a reason. He is beyond the *reach* of such reasons, at least as things currently stand. We are consigned to doing something more in the vein of giving him an order, or simply expressing our disapproval. It can be important for us to state 'for the record' that we find both his actions and his attitude unacceptable. We may also have to resort to trying to manage this man's behavior—by sending him to anger management class, or helping his wife to get out of there, or assisting her in obtaining a restraining order against him. Or we may have to try to get him arrested, hoping that he will be locked up, at least until he can be reformed and hopefully rehabilitated. Any of these interventions might well be the sort of thing which we ourselves have good or decisive reasons to do. Our hands are not

Footnote 50 continued

people like the callous husband—who are plausibly autonomous in the standard sense but morally or socially 'out of it'—with people like the 'crazy' or intoxicated person—whose autonomy *is* compromised—in certain of our moral-cum-social interactions with them.

tied. But this man's motivational profile should make a fundamental difference to our sense of where we stand in our moral-cum-social relationship to him. Our stance towards him is (or, at least, should be) no longer interpersonal, at least to the extent that we are still trying to influence the way he treats his wife.⁵¹ Rather, it becomes (or, again, should become) objective. In this sense, our tongues are tied by his motivational deficits. This is admittedly sad, but it may be true nonetheless.

Why should it be sad, though? For, of course, I do not mean to suggest that anyone on the opposing side thinks that there is always a realistic possibility of convincing the callous husband and his ilk to do otherwise on the basis of offering him certain supposed reasons. Nobody is *that* optimistic, I take it. Still, I am especially gloomy in the following respects.⁵² Firstly, I believe that it is not merely pointless but in a certain way senseless to suggest in a reasonable tone of voice that such people do otherwise, given the limitations of their current motivational repertoire. Some theorists seem to think, in contrast to this, that we are obliged or at least entitled to say that these people should behave better, unless they are positively arational in some way. Even if we won't *get* anywhere, we can still take a moral stand—a moral stand of the very same sort as we were always going to take here. More abstractly, some theorists think that what is wrong with the callous husband and his ilk is a failure of some form of rationality or reasonability, such that he is morally little different from a well-disposed person who is being obtuse or inconsiderate, but could be made to see the light. Similarly, certain of my opponents seem to believe that there is a difference in degree rather than in kind between the sort of moral criticism which applies to someone who is sometimes lazy about recycling his newspapers and whatnot, and somebody who really does not care about the future of the planet. They are both failing, albeit to different extents, to act on reasons which they have—which are somehow in their possession just as reasonable creatures.⁵³ I find this sort of rationalism difficult to accept, and I hope to have now offered some theoretical support for this intuitive discomfort. For, as I've been arguing, I think the limitations of the interpersonal activity of reasoning with someone are well-drawn at the boundary of the motivational propensities of the person we're trying to get through to.

I would also resist those who might claim (alternatively or additionally) that it is one thing to admit that we shouldn't or indeed couldn't *say* certain things to the callous husband, in our capacity as his ideal advisor, but another thing altogether to

⁵¹ Note that the suspension of the interpersonal stance towards someone can be quite local. Someone may be impossible where women but not men are concerned, for example. This is one form or symptom of a misogynistic moral 'blind spot.'

⁵² Thanks to Alan Kim and Timothy Rosenkoetter for pressing me about this—and also to Dennis Whitcomb for trying to lift my mood here.

⁵³ And this continuity is meant to matter here. Parfit: "It matters greatly, I believe, whether Hitler had reasons to do what he did." (2011, vol. 1, p. 99) And compare Gilbert Harman's well-known remark that "It sounds odd to say that Hitler should not have ordered the extermination of the Jews, that it was wrong of him to have done so. That sounds somehow "too weak" a thing to say. Instead we want to say that Hitler was an evil man." (1975, p. 7) I mean to be saying that there is something similarly off-key about reasons-talk as applied to even the little Hitlers of the world.

deny that they are true.⁵⁴ From the practice-based perspective, a distinction of this kind is not supposed to be sustainable. Of course, normative considerations can hold despite the pragmatic reasons not to offer them to an agent on a particular occasion, or given the ways in which our actual conversation falls short of some ideal. But it simply does not make sense to suppose that a normative reason for action could still apply to an agent despite the fact that there would never be any point in offering this consideration *to* her, within the practice of reasoning with her about what she ought to do. Insisting that this is possible is thus to beg the question against the practice-based approach to normativity which I've been trying to bring into view here. And I think that this approach may be attractive albeit confronting on a number of different levels. It appeals to me in particular as someone who thinks of practical authority as primarily being a property of people and not of facts. Hence my sense that the practice-based approach to normativity is worth at least exploring further, together with its implications.

3 You ought to have a reason...

In the first section of this paper, I recommended that we accept that reasons for an agent *A* to ϕ are considerations which would be apt to be cited in favor of *A*'s ϕ -ing, within the activity of reasoning with her about what she ought to do, on the part of her ideal advisor. In the second section of this paper, I argued that citing some consideration in favor of *A*'s ϕ -ing is only going to be apt within the context of reasoning with *A* about what she ought to do if she would become (somewhat) motivated to ϕ , following an idealized process of being reasoned with in this way. It follows that there is the corresponding motivational constraint on an agent's reasons for action: namely, a reason for an agent *A* to ϕ can hold only if *A* would become (somewhat) motivated to ϕ , following an idealized process of being reasoned with in this way.

Here's the obvious objection, though, to all of what has come before: as well as various practical interventions of the kind mentioned above, don't we want to be able to criticize or even roundly condemn people like the callous husband? Don't we need to be able to point the finger at him with impunity, if we are to have any hope of holding onto our moral hats here? In particular, mustn't we reserve the right to *blame* him for not being nicer to his wife? For, blame might be thought to be a paradigmatic reactive attitude. It therefore seems plausible to think that we will have to give it up, in abandoning the interpersonal stance towards him with regard to the matter in question.⁵⁵

⁵⁴ Compare the line taken by Scanlon in his insightful discussion of reasons internalism (1998, Appendix). And thanks to Alex Guerrero for pushing me on this point.

⁵⁵ Strawson leaves off talking about blame specifically, following his general opening remarks in "Freedom and Resentment" about our practices of punishing and blaming. But I take it he would have agreed here. Williams might not have. In his well-known paper on the subject of blame, Williams distinguishes briefly between the sort of 'focused' blame that tracks an agent's reasons and "hopes to achieve recognition," and a less committal form of blame which merely represents "a rejection (perhaps an entirely justified rejection) of what [the agent] has done." (1995a, p. 44) But elsewhere Williams

I think that there are two things to say in response to these crucial worries. The first is that there are other critical things we can say about this man and his behavior. We have given up a specific kind of moral criticism (remembering that we are provisionally rejecting the idea that reasons-talk can capture everything that we might want to say in a critical tone of voice). But the kind of moral criticism which we are contemplating giving up here is undeniably important. The blame game matters. This is where I think that confusions can arise, though. We may mistake a confronting thesis for one that is counterintuitive.

Let me canvass the first, more conciliatory point to begin with, and come back to the second, more confronting point in closing. When we give up saying that the callous husband and his ilk have reasons to do otherwise, what might we say instead, along broadly critical lines? As I have already recorded, I think we can criticize this man's actions, by saying that he is acting badly, or evaluate his behavior in some other negative way. It would certainly be good if he were to develop different motives. (Williams: "Things, and he, would be a lot nicer if he were so motivated;" 1981a, p. 110) For, it is a bad state of affairs that his wife is being treated in a callous way. (Williams: "I shall presumably say, whatever else I say, that it would be better if he were nicer to her;" 1995a, p. 39.) And as well as saying these various things, we may be entitled to do various things to try to manage or otherwise influence this man's bad behavior, as I've also already recorded. Notably, some of these strategies will make essential recourse to his being in some ways perfectly reasonable. For example, we may offer him incentives or issue him with threats which tap into motivations he does have—thus *giving* him reasons to do things which he did not have before.⁵⁶ And there is also the possibility of managing this man's behavior by saying certain things which are not strictly speaking true, as Williams himself emphasized. (1995a) Suppose, for example, that he is anxious to secure our good opinion of him. Then we might say "There's a reason to be nicer to your wife!" in the objective mode, hoping that this (currently false) statement will act as a kind of blunt instrument. That is, we might hope that, in saying this, we will somehow change this man into a person who is appropriately receptive to our moral remonstrations—on account of his awareness that this is the expectation. How and when this might work more precisely are questions for another day, though.

Williams pointed these things out, but his opponents have not been greatly impressed. Often, in conversation, I've found that people tend to emphasize the need to say something critical about the man himself. Williams did acknowledge

Footnote 55 continued

construes our ordinary notion of blame as belonging with reasons-talk; see *n.* 24. And the sort of putative blame that involves mere hostility and rejection might be better termed condemnation.

⁵⁶ Tapping into people's existing motivations in such ways may but need not involve our taking leave of the interpersonal stance towards them. We often aim to influence people's behavior by making certain requests of them, or having them make us promises, rather than by reasoning with them. Someone might ask his spouse to promise to look after herself properly while he is away travelling, thereby giving her a moral reason to do things she doesn't care about doing for her own sake, such as eating regular meals. But she might care about being true to her word—hence the point of extracting a promise of this kind from her.

this, if perhaps a little blithely. He wrote that “there are many things I can say about or to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things.” (1995a, p. 39) Still, we might wonder how to take these criticisms, if they are not bolstered by an underlying reasons-claim. Are they anything more than name-calling? Parfit especially has doubts.⁵⁷

I think it is fair enough to want to secure a connection between criticisms of this man and the reasons-claims which we’d hoped to be able to make on him (but which the internalist says we cannot, in view of his motivational deficits). Still, we have to be careful about assuming that this must be a simple connection. In other domains of inquiry, not every criticism of a person can be well-expressed by talking about the reasons they failed to comply with. For example, not every epistemic flaw can naturally be construed as someone failing to believe in accordance with their evidence, or what they have reasons to believe. They may also exhibit broadly epistemic flaws insofar as they fail to gather any evidence, and so fail to acquire certain reasons to believe which they ought properly to have. Or at least, this is one way of getting at a distinction which is natural and important.

Examples which are structurally similar in some ways can occur in legal cases too. Suppose that there ought to be a legal obligation to register one’s dog with the city of Somerville council (to take a prosaic example adapted from Ruth Barcan Marcus 1966). And if there were to be such a law, I would then have a reason to attach a registration tag to my dog Panko’s collar. (This is the way such laws standardly work, since their purpose is partly to help differentiate lost pet dogs from strays.) But there is no such law. Do I have a reason to attach a City of Somerville registration tag to my dog Panko’s collar? Well, no: how could I? I not only lack such a tag, but these tags do not even exist. I could not procure one with the right sort of legal status if I tried. So I ought to have a reason which I currently and regrettably lack.

I do not want to suggest that either of these analogies is perfect. Still, I think they are suggestive of the following possibility. Perhaps the problem with the callous husband—the sense in which he is a morally and socially flawed person—is not that he fails to act on reasons which he has, but rather that he fails to even make it to base camp. That is, the problem is precisely that he lacks certain reasons which ought to be had by everyone (or at least everyone who’s married). So, again, he ought to have a reason to be nicer to his wife. But, sad to say, he doesn’t.⁵⁸

There are various complexities which the analogies rightly suggest here. For example, notice that an agent’s failure to gather certain evidence might lead us to

⁵⁷ Parfit: “When cruel people make others suffer, we can call these people vicious, odious, and callous. But on subjective desire-based theories, some of these people have no reason not to make others suffer. These other criticisms become much weaker if we must admit that, on our view, these people have no reason to act differently.” (2011, vol. 2, pp. 456–457)

⁵⁸ Compare Williams’ remarks about the “deficiency or fault” of a man who does not recognize certain moral considerations. Williams agrees that “certainly he cannot head off the criticism by saying that the reasons do not apply to him because he does not have that kind of [subjective motivational set].” But “none of this implies that these considerations are already the defective agent’s reasons; indeed, the problem is precisely that they are not.” (2001, p. 96)

criticize the agent herself, but arguably only if she failed to act on reasons she had to gather more evidence in the first place. In contrast, in the legal case, whatever criticism there might be should obviously be directed towards other people (i.e., those responsible for making the law). Moral cases might take either form, for all that has been said so far. And on the face of it, they take both. On the one hand, there may be instances in which we regard the person as having ‘let himself go,’ morally, or ignored the reasons he had for cultivating or retaining his moral sensibility (say). In such cases, blame in the standard sense may be quite appropriate. But it may be blame for his past negligence—much as we might blame a person who was driving drunk not for falling asleep at the wheel, nor even for getting into the car when he was well past the point of being able to make a rational decision. Rather, we primarily blame him for not handing over his keys at the beginning of the evening, and then drinking to excess, knowing that this could happen. In other cases, blame in the standard sense may be off the table entirely—instead being redirected, quite properly, at other people who have effectively made the agent into the “dangerous or hopeless” character who we see before us today. (1995a, p. 43) In still other cases, there may be nobody to blame. And this can obviously be sad and also bewildering in itself. Why are some people so indifferent towards others?

This brings me to the next and final point I want to make here. It may be not only sad but frightening that a person lacks moral reasons which he ought to have in this sense—especially when others’ well-being hangs in the balance of his doing what he has no reason to do, according to the internalist. Deep break-downs in our moral-cum-social relations are not a happy thing to behold. And essentially unguided missiles who come in the shape of people are all the more unnerving and dispiriting to deal with, since the reminder of what—or, rather, who—could have been is staring you right in the face. There would thus be something reassuring about the thought that everyone who is not positively arational, in being (e.g.) mad, drunk, or very young, can be kept out of this category in our moral-cum-social relations. The thought essentially then being that we who are rational agents are all playing the same moral game, even if some of us play it badly. But I think that there is no real basis for believing as much, unfortunately. There are obviously empirical questions here, along with moral and conceptual questions about where reasoning with someone stops and other practices begin. Still, one has one’s hunches. And it seems to me that there are some people who are most plausibly interpreted as being beyond the reach of being reasoned with about what they are doing. Most abusive husbands are not like this—many want to change quite desperately—but there is the occasional one who is (or so domestic violence victim advocates have told me, if I am extrapolating from what they say correctly).⁵⁹ And many people can be talked into getting interested in what they can do in the fight against global warming, if we are really persistent. But some people probably cannot, and are seemingly not even willing to have the conversation. They really do not give a damn about the quality of

⁵⁹ Thanks to my own wonderful husband, Daniel Manne, a legal academic whose research centers on domestic violence, for helping me not to lose sight of the complicated realities here, which affect the lives of so many women (in particular).

life that will be enjoyed (or not, as the case may be) by future generations. Williams spoke of our feelings of “rage, frustration, sorrow, and fear” in these sorts of situations (1981b, p. 122). And we still feel the need to say something about and perhaps also to these people, I freely admit. We may even still feel the need to say that they have *reasons* to do otherwise, notwithstanding the other possibilities which I’ve pointed to in this section. But we should be careful of mistaking a psychological need of ours for a philosophical certainty, or even an intuition. As Williams memorably put it, we desperately want reasons-claims to “‘stick’ to the agent; but the only glue for this purpose is social and psychological.” (1981b, p. 122) This glue is not always strong enough; it may fail to get a grip entirely. Hence my pessimistic conclusion: internalism about reasons may represent the sort of sad truth about people we just have to learn to live with.

Acknowledgments Earlier versions of this paper were presented at a colloquium at the University of Pittsburgh, a class at Dartmouth College, a meeting of WOGAP at MIT, and the Bellingham Summer Philosophy Conference (BSPC) in 2013. Section 3 contains material originally presented at Washington University in St. Louis and Cornell University in 2011, and also in the Young Philosophers’ Lecture Series held at SUNY Fredonia in 2012, thanks to Andrew Cullison. I’m grateful to all of these audiences for their helpful questions and comments on my developing ideas here. And I’m grateful to the BSPC organizers—Julia Markovits, Miriam Schoenfield, and Ned Markosian—for all of their hard work in organizing this fabulous occasion. In terms of the substance of this paper, I’d also like to thank the members of my dissertation committee at MIT—Richard Holton, Sally Haslanger, Julia Markovits, and Rae Langton—for their invaluable feedback on Chapter 2 of my dissertation, which this paper essentially grew out of. Further thanks to Kenneth Walden and Kieran Setiya for recent fruitful discussions, and also to Tyler Doggett and Hille Paakkunainen for very generous and valuable sets of written comments which they were kind enough to send me. Finally, I’m indebted to Julia Driver and Alex Guerrero for their terrific commentaries on this paper at BSPC, both of which helped me a great deal. I’ve had occasion to thank several others along the way who have helped me in thinking through various specific issues here. But I am sure I am forgetting people who were also kind enough to share in my sadness—or, alternatively, try to cheer me up a bit.

References

- Barcan Marcus, R. (1966). Iterated deontic modalities. *Mind*, 75, 580–582.
- Finlay, S., & Schroeder, M. (2012). Reasons for action: Internal vs. external. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2012 Edition). <http://plato.stanford.edu/archives/win2012/entries/reasons-internal-external/>.
- Harman, G. (1975). Moral relativism defended. *Philosophical Review*, 84(1), 3–22.
- Manne, K. (2013). On being social in metaethics. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 8, pp. 50–73). Oxford: Oxford University Press.
- Manne, K. (2014). Tempered internalism and the participatory stance. In G. Björnsson, C. Strandberg, R. F. Olinder, J. Eriksson, & F. Björklund (Eds.), *Motivational internalism*. Oxford: Oxford University Press.
- Markovits, J. (2014). *Moral reason*. Oxford: Oxford University Press.
- McDowell, J. (1998). Might there be external reasons? In *Mind, value, and reality* (pp. 65–85). Cambridge, MA: Harvard University Press.
- Parfit, D. (2011). *On what matters* (Vol. 1 and 2). Oxford: Oxford University Press.
- Railton, P. (1986). Facts and values. *Philosophical Topics*, 14(2), 5–31.
- Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Schroeder, M. (2007). *Slaves of the passions*. New York: Oxford University Press.
- Smith, M. (1994). *The moral problem*. Oxford: Blackwell.
- Sobel, D. (2001). Explanation, internalism, and reasons for action. *Social philosophy and policy*, 18(2), 218–235.

- Strawson, P. F. (1962/2008). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25. (Reprinted in *Freedom and resentment and other essays*, 2nd ed., pp. 1–28, New York: Routledge).
- Velleman, J. D. (ms). Time for action.
- Williams, B. (1981a). Internal and external reasons. In *Moral Luck* (pp. 101–113). Cambridge, MA: Cambridge University Press.
- Williams, B. (1981b). Ought and moral obligation. In *Moral Luck* (pp. 114–123). Cambridge, MA: Cambridge University Press.
- Williams, B. (1981c). Practical necessity. In *Moral Luck* (pp. 124–131). Cambridge, MA: Cambridge University Press.
- Williams, B. (1985). *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.
- Williams, B. (1995a). Internal reasons and the obscurity of blame. In *Making sense of humanity* (pp. 35–45). Cambridge, MA: Cambridge University Press.
- Williams, B. (1995b). Replies. In J. E. J. Altham & R. Harrison (Eds.), *World, mind, and ethics: Essays on the ethical philosophy of Bernard Williams* (pp. 185–224). Cambridge, MA: Cambridge University Press.
- Williams, B. (1996). Values, reasons, and the theory of persuasion. In *Philosophy as a humanistic discipline* (pp. 109–118). Princeton, NJ: Princeton University Press.
- Williams, B. (2001). Postscript: Some further notes on internal and external reasons. In E. Millgram (Ed.), *Varieties of practical reasoning* (pp. 91–97). Cambridge, MA: MIT Press.