RECENT WORK ON INTERNAL AND EXTERNAL REASONS[1]

John Brunero

Abstract

The paper examines some recent arguments for internalism which (i) appeal to an analogy between practical and theoretical reasons, (ii) look toward our practices of reasoning with others, or (iii) tie reasons to good deliberation. The conclusion of this paper is a skeptical one: none of these new arguments give us sufficient reason to think that internalism is true.

In "Internal and External Reasons," and in later, related papers, Bernard Williams argues that an agent's reasons depend on his psychological makeup, specifically on the contents of his "subjective motivational set" – a set which includes his desires, projects, loyalties, and commitments, as well as his "dispositions of evaluation" and "patterns of emotional reaction."[2] More precisely, Williams argues that claims about there being, or one's having, reasons to ϕ – claims of the form "There is a reason for A to ϕ" or "A has a reason to ϕ" – are true only if there is a "sound deliberative route" from A's subjective motivational set to the conclusion to ϕ (or the conclusion that A should ϕ).[3] In Williams's terminology, an "internalist" view is one which accepts this necessary condition, while an "externalist" view is one which denies it.[4] Williams was somewhat liberal about what counts as sound deliberation, allowing that it could involve more than deliberation about how best to satisfy one's desires, and could include, for instance, deliberation about how much weight to attach to parts of one's subjective motivational set in cases where they conflict with one another.[5] Despite that liberality, this necessary condition for the existence of a reason constitutes a significant, and perhaps counterintuitive, constraint on what agents can have a reason to do. To see this, consider the example Williams presents of a nasty husband who has no motivations in his subjective motivational set that could be

connected, via sound deliberation, to the conclusion to be nice to his wife.[6] In that case, Williams argues, we must concede that there is no reason for him to be nice to his wife.

The application of Williams's internalism to examples like this one – examples where moral obligations seem applicable – explains the importance of his view to meta-ethical theorizing. On Williams's view, there is no such thing as a *categorical reason* – that is, a reason for someone to do something regardless of what he desires. But morality, on a fairly ordinary understanding of it, seems committed to the existence of such reasons.[7] Thus, Williams's view seems to present a challenge to that ordinary understanding of morality. And, naturally, philosophers have appealed to internalism in support of such meta-ethical views as an error theory about morality[8], moral relativism[9], and moral anti-rationalism.[10] However, other philosophers have suggested that the appearance of a conflict between an internalist theory of reasons and our ordinary understanding of morality might be illusory.[11]

Given the importance of Williams's view to metaethical theorizing, there is a large literature devoted to reconstructing Williams's argument for his view, which, very roughly, appeals to a conceptual connection between reasons and the explanation of action.[12] However, I think it's fair to say that Williams's argument is somewhat unclear, and few of those philosophers who claim to understand it are convinced by it.[13] If Williams's own argument isn't clear and convincing, we are left with the question of why we should think internalism is true, especially if it issues counterintuitive predictions, as in the case of the nasty husband. Fortunately, there has been a renewed interest in the prospects for Humean theories of reasons in recent years, which has produced several new arguments for internalism.[14] The aim of this paper is to consider and assess some of those arguments.

Rather than aim for comprehensive coverage of recent work on internal and external reasons, I will here discuss three important trends in recent work on this topic. In particular, I'll focus on arguments for internalism which (§ 1) appeal to an analogy between practical and theoretical reasons, (§ 2) look toward our practices of reasoning with others, and (§ 3) tie reasons to good deliberation. The conclusion of this paper, however, will be a skeptical one: none of these new arguments give us sufficient reason to think that internalism is true.

1. THEORETICAL REASONS

One might expect that our philosophical accounts of reasons for action (practical reasons) and of reasons for belief (theoretical reasons) shouldn't yield the result that the two kinds of reasons are radically different. After all, it's plausible to think that there's just one concept here, the concept of a reason, applied to both actions and beliefs. On this assumption, it's also plausible to think that an investigation into the nature of theoretical reasons could help settle disputed questions about the nature of practical reasons, including the question of whether internalism is true. In this section, I'll consider two ways in which this argumentative strategy has been employed, one due to Alan Goldman (2009) and another due to Julia Markovits (2014).

The idea that this argumentative strategy would support *internalism* is surprising. At first glance, it seems to support externalism, since it's commonly thought that theoretical reasons do not depend for their existence on an agent's desires.[15] If you are presented with evidence for some proposition, that evidence provides you with a reason to believe the proposition, even if you have no desires that would be promoted by your doing so. You may even hinder the promotion of certain desires by forming true beliefs

on some subject. For instance, perhaps you want to avoid learning how some movie ends.[16] Forming true beliefs about how it ends does not promote any, and indeed hinders the promotion of some, of your desires. Yet the provision of reliable testimony that it ends tragically nonetheless provides you with a reason to believe it ends tragically. Since reasons for belief seem categorical – in that their existence doesn't depend upon an agent's goals or desires – it seems that the practical-theoretical analogy would support externalism, not internalism.

Goldman, however, suggests that the practical-theoretical analogy provides "a powerful argument" for internalism if we look to the *constitutive aims* of action and belief.[17] It's often noted that the constitutive aim of belief is truth. And some philosophers have thought that this constitutive aim sets a standard for successful belief (a successful belief is one which is true) and also determines what counts as reasons for belief (reasons for belief are the *indicators* of truth, namely the evidence).[18] Since this aim is a *constitutiv*e aim, it partly defines what it is for something *to be* a belief: if some psychological state fails to be so aimed, it fails to count as a belief.

It's worth pausing to note that the idea that belief's constitutive aim determines what counts as reasons for belief need not be in tension with the idea, mentioned above, that theoretical reasons are categorical. That's because talk of a constitutive "aim" is largely metaphorical, or at least need not be identified with an *aim of an agent*. As David Velleman puts it, "Truth must be the aim of belief, but it need not be an aim on the part of the believer; it may instead be an aim implicit in some parts of his cognitive architecture."[19] In other words, truth may be identified as a "sub-agential," rather than agential, aim. If it is so identified, one could hold that theoretical reasons are determined

by the constitutive aim of belief while also maintaining that these reasons are categorical in that they do not depend for their existence on an *agent's* goals or desires.

Assuming that the constitutive aim of belief sets a standard for successful belief (truth) and determines what counts as reasons for belief (indicators of truth), we can then search for the constitutive aim of action, which, likewise, would set a standard for successful action and determine what counts as reasons for action. So, the next step in Goldman's argument is to identify the constitutive aim of action, which he takes to be the satisfaction of the desires that prompt an action: "Philosophers seem to miss this condition or constitutive aim because it is right before their eyes. Actions are successful when they fulfill the motivations that prompt the actions."[20] And reasons for action "indicate ways or means of succeeding" in the satisfaction of those desires.[21] Goldman takes this to constitute "a strong argument for internalism, the thesis that all our practical reasons derive from our concerns or motivations."[22]

But is Goldman correct that the constitutive aim of action is the satisfaction of the desires that prompt an action? A constitutive aim of action is supposed to be such that if some behavior lacks that aim, it doesn't count as an action (just as if some attitude isn't aimed toward truth, it doesn't count as belief). One might worry that not all actions – and perhaps, more strongly, not all intentional actions – are prompted by some desire which they aim to satisfy. For instance, this morning I intentionally kicked my garage door out of frustration with a household repair gone wrong, knowingly bringing about pain in my foot and a dent in my door, neither of which were desired outcomes. However, it's open to Goldman to insist that there was some desire that prompted the action – perhaps a desire to kick the garage door – and that my bodily movements were

in some sense aimed at bringing about the satisfaction of that desire. Let's assume that such a line of reply is acceptable.

But the worry is that it's now unclear how this alleged constitutive aim of action yields a plausible theory of normative reasons. While Goldman's view does allow for the possibility of unsuccessful actions – I might whiff in my attempt to kick the door, for instance, thereby not satisfying my desire to kick the door – one might worry that success is too easy to come by. My kicking the door out of frustration (when I don't whiff) is a fully successful action, since it satisfies the desire that prompted the action. It is no less and nor more successful than my action of walking away would have been, had my doing so been prompted by a desire to walk away. But, of course, we want to deliver the verdict that there is *more reason* for me to walk away than there is for me to kick the door. But if doing either of these things would *equally well* satisfy the desires which prompt them, thereby equally well achieving the constitutive aim of action, it's unclear how we can deliver this result.

Goldman seems aware of this difficulty and offers a response. The response would allow him to say that an agent who satisfies his desire to kick the door, thereby perhaps hindering the promotion of his "deeper" desire to avoid pain, is guilty of practical incoherence. He writes:

> Likewise, it might be claimed, in desiring p an agent aims to satisfy only that desire, not her deepest desires at the time, as I claim rationality requires. But the aim of truth also generates a requirement of coherence for beliefs, as I have argued. It defeats that aim to believe inconsistent propositions. Likewise, the aim of maximal desire satisfaction requires practical coherence. Rational agents aim

> first to satisfy their deepest desires occurrent at the time, since it is self-defeating to satisfy a weaker or shallower desire at the expense of a stronger or deeper one.[23]

However, this reply shifts the constitutive aim of action. Initially, it was *satisfying the desires that prompted the action*. But this passage refers to the "aim of maximal desire satisfaction," which generates a requirement of coherence prohibiting the satisfaction of "shallow" desires at the expense of "deep" ones. And this aim of maximal desire satisfaction is now presented as analogous to belief's aim of truth. The view has shifted.

So, we should consider whether maximal desire satisfaction is a plausible candidate for action's constitutive aim. Although full discussion of this question would require more space than we have here, there's good reason to be skeptical. As Velleman observes, the standard of correctness for belief is *uncontroversial*, since it's just part of the very concept of belief that it's an attitude that can be correct or incorrect by being true or false.[24] But Goldman's account of the standard of correctness for action isn't similarly uncontroversial. It's hard to think that it's part of the very concept of action that an action is correct just when it maximally satisfies one's desires. (Moreover, an externalist might respond simply by insisting the standard of correctness is the maximal promotion of *value*, or something similar. It's not clear what rules this view out.) Additionally, while it might have been plausible to think that my action of kicking the garage door is aimed at the satisfaction of the desire to kick the garage door, it's harder to see in what sense that action is aimed at *maximal desire satisfaction*. Specifically, it's hard to see how in kicking the door in front of me, I am attempting to maximally satisfy my desires *in the same sense* in which in believing there's a door in front of me, I'm aiming to correctly represent things as they really are.

In summary, Goldman's appeal to the practical-theoretical analogy looks to the constitutive aims of action and belief. But, on either of the suggested accounts of the constitutive aim of action, there are difficulties. If the constitutive aim of action is taken to be the satisfaction of the desires that prompt an action, Goldman's view doesn't yield an intuitively correct account of the reasons we have. But if the constitutive aim of action is taken to be maximal desire satisfaction, it's less clear that this aim is constitutive of action in the same way that truth is the aim constitutive of belief.

Julia Markovits (2014) also develops an argument for internalism based on the practical-theoretical analogy. She argues that just as your reasons for belief, for the most part, depend only upon what you already believe and the standards of procedural rationality, your reasons for action depend only upon what you already desire and the standards of procedural rationality. (I'll explain why the "for the most part" qualification is necessary in the theoretical, but not the practical, case in a moment.) Thus, if we want an account of practical reasons analogous to theoretical reasons, internalism is the way to go.

As Markovits understands it, internalism about theoretical reasons is the view that "what we have reason to believe depends *only* on what we already believe and the standards of procedural rationality."[25] However, she thinks that internalism about theoretical reasons is false, since some theoretical reasons – namely, reasons provided by sensory experience – do not depend only on what we already believe and the standards of procedural rationality. (That explains the "for the most part" qualification above.) For instance, the fact that I'm experiencing redness gives me a reason to believe I'm experiencing redness, regardless of my antecedent beliefs. However, there is nothing analogous to the reasons provided by sensory experience in the practical case. (The

practical counterpart of such reasons would have to involve "ends that are uncontroversial, largely immune from erroneous adoption, and therefore not the kinds of things we feel people must offer further justification for caring about, beyond telling us they care about them."[26] And there are no such ends.) Since there is nothing analogous to these "external" theoretical reasons in the practical case, we can safely conclude that the practical-theoretical analogy supports internalism about practical reasons.

One way to resist Marokovits's argument is to challenge her (largely) belief-dependent conception of theoretical reasons. On that conception, one's reasons for belief, except for those provided by sensory experience, depend only upon what one already believes and the standards of procedural rationality; if some proposition lacks a rational connection to one's current stock of beliefs, there's no reason to believe it. Suppose, for instance, that a patient has a severe allergy to some drug, and a physician, not knowing about this allergy, forms the belief that the drug will make the patient better off. [27] But the physician is wrong, and giving the drug to the patient will only make things much worse. On Markovits's view, if there's no rational connection between the physician's current stock of beliefs and the proposition that the drug will make the patient worse off, then she has no reason to believe that proposition. But Markovits's view is controversial. Suppose that the explanation for the lack of such a rational connection is that the physician's current stock of beliefs is impoverished due to her negligence. Perhaps she didn't ask about the patient's allergies or bother to look carefully at the medical chart. It's clear that this makes her guilty of medical malpractice. But it is also natural to think that there were conclusive reasons for her to believe the drug would make the patient worse off, and, because of her insensitivity to those reasons, she failed to

believe what she ought to have believed: that the drug would make the patient worse off. She should have known better.[28]

In light of examples like this one, it's natural to think that reasons for belief are at least sometimes provided by the *available evidence*, and so aren't entirely constrained by an agent's current stock of beliefs and the standards of procedural rationality.[29] There's a question about how to understand availability, but regardless of how we do that, we've here further severed the tie between a person's theoretical reasons and her current stock of beliefs, thereby undercutting the support that the practical-theoretical analogy would provide for internalism about practical reasons.

Another way to resist Markovits's argument is to argue that the motivations for making theoretical reasons belief-dependent do not transfer over to making practical reasons desire-dependent. Consider a variation on the physician case in which the physician isn't negligent, but is instead excessively thorough, yet the patient nonetheless has an extremely rare, unexpected allergic reaction to the drug. It seems odd to say here that the physician had a reason to believe the drug would make the patient worse off. After all, she was thorough, and reasoned impeccably from the evidence available to her. She's *not an appropriate target for blame or criticism*. And that's a central motivation for making her theoretical reasons depend upon her current stock of beliefs, the available evidence, or something similar (as opposed to making them depend on inaccessible facts about the patient's allergy).[30]

But does this central motivation for thinking theoretical reasons depend on beliefs also support the thesis that practical reasons depend on desires? It doesn't seem so. One doesn't get off the hook when it comes to blame and criticism simply by *not caring*. If we suppose the physician knows the drug will provoke an allergic reaction in the patient, but

has no desires that would be promoted by not prescribing it, and she prescribes it, she's still subject to blame and criticism.[31] So the central motivation for making theoretical reasons belief-dependent doesn't also carry over to support making practical reasons desire-dependent. And that should make us suspicious of the thought that practical reasons are analogous to theoretical reasons in the way Markovits suggests.

In summary, Markovits develops an argument for internalism that appeals to the practical-theoretical analogy, suggesting that just as theoretical reasons (for the most part) depend on a person's antecedent beliefs, practical reasons depend upon a person's antecedent desires. But it's not clear that she provides the correct account of theoretical reasons. And even if she does, since the central motivation for the belief-dependency of theoretical reasons doesn't also support desire-dependency in the practical case, we have principled grounds for resisting the analogical argument for internalism.

2. REASONING WITH OTHERS

Another interesting argument for internalism, developed by Kate Manne (2014), appeals to the relationship between reasons and the social practice of *reasoning with others*. Manne argues that some consideration is a reason for an agent to ϕ only if that agent would come to recognize that consideration as a reason after her ideal advisor *reasons with* her. If that's right, and if we also endorse a version of judgment internalism according to which the recognition of a consideration as a reason to ϕ must involve a motivation to ϕ[32], we can immediately derive a version of internalism: a person has a reason to ϕ only if she would be motivated to ϕ after her ideal advisor reasons with her.[33]

But why should we accept the first premise here – that a person has a reason to ϕ only if she would come to recognize that consideration as a reason after her ideal advisor reasons with her? Manne's argument for this claim rests on a view about the connection between reasons and *what an ideal advisor would be apt to cite*:

> 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."[34]

The ideal advisor is described as fully informed, procedurally rational, and well-disposed toward the advisee (and perhaps also "virtuous and wise").[35] Most importantly, the ideal advisor is someone who is in the best position to play the social role of reasoning with A; she is "the person who is best suited to 'getting through' to her, morally."[36] This social role is also characterized by Manne as that of "reasoning with her, or (similarly) offering her collaborative advice or friendly suggestions, about what she ought to do."[37]

When the ideal advisor is playing this social role of reasoning with the advisee – at least when she is doing so in an ideal way[38] – she adopts what P.F. Strawson calls "the interpersonal stance" towards the advisee – that is, she treats the advisee, as Manne puts it, "as a sovereign creature … both equipped and entitled to make [her] own decisions," and over whom she has no authority.[39] (This contrasts with taking Strawson's "objective stance" toward the advisee, in which one might view the advisee, as Manne puts it, "as a kind of human object to be managed, cured, or navigated around."[40]) Manne then argues that when an advisor doesn't pay attention to the motivational profile of the advisee – specifically, when she attempts "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" – she no longer adopts the interpersonal stance towards her.[41] Since the ideal advisor is fully informed, she will be perfectly familiar with the motivational profile of the advisee, and so will know in advance which considerations would be recognized as reasons and which would not. Given that knowledge, the advisor will not – at least insofar as she is *reasoning with* the advisee, which, again, involves taking the interpersonal stance towards her – cite some consideration as a reason that the advisee would not accept as a reason.

We can now put these two pieces together. If an agent's reasons are just those considerations an ideal advisor would be apt to cite when reasoning with the agent, and an ideal advisor reasoning with an agent would be apt to cite just those considerations that the agent would come to recognize as reasons, then an agent's reasons are just those considerations she would come to recognize as reasons after her ideal advisor reasons with her. That completes Manne's argument for the first premise above.

It's worth noting that Manne's internalism is in one way *more restrictive* than Williams's internalism. For Williams, a reason to ϕ exists only if there's a sound deliberative route from the agent's subjective motivational set to the conclusion to ϕ. Suppose some alleged reason meets Williams's condition, but is nonetheless such that the agent wouldn't, even after an ideal conversation with an ideal advisor, recognize it as a reason. In that case, Manne would disagree with Williams and hold that this consideration isn't a reason.[42]

How will externalists (and perhaps also fans of Williams's internalism) reply to Manne? A natural line of response would be to insist on a distinction between the *truth* and the *rhetorical effectiveness* of reasons claims. An externalist, for instance, would insist

that in the case of the nasty husband, it's true that he has a reason to be nice to his wife – perhaps her needs and interests constitute such a reason – but this reasons claim is rhetorically ineffective, since it wouldn't be recognized by the husband as a reason, even after conversation with an ideal advisor. Additionally, the externalist would likely appeal to conversational pragmatics to explain why an ideal advisor wouldn't be apt to cite such true, but rhetorically ineffective, reasons claims when reasoning with her advisee. (Perhaps doing so would be a violation of Grice's Cooperative Principle, since the mutual conversational ends make the contribution of that reasons claim irrelevant, or perhaps doing so would be a violation of some norm of *politeness*, since, if you're looking to avoid discord, you'd leave it be, rather than press some consideration that you *know for certain in advance* won't get any uptake.[43]) However we explain it, what's crucial for the externalist (or the fan of Williams's internalism) is the possibility of true, but rhetorically ineffective (even in ideal conditions) reasons claims.

However, this line of reply begs the question against Manne's view, since it's part of Manne's "practice-based" approach to normativity that the truth of reasons claims isn't a separate matter from their rhetorical effectiveness: "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."[44] Is there a way to resist Manne's view of reasons without begging the question against the practice-based approach?

One strategy, discussed in the previous section, would be to consider the plausibility of Manne's view for theoretical reasons. It's commonly thought that one's reasons for belief are provided by the available evidence. But suppose that we consider a

person who is so dogmatic, brainwashed by propaganda, or prone to wishful thinking on some subject that she isn't receptive to the evidence presented to her, and wouldn't be receptive even after being reasoned with by an ideal advisor. If we apply Manne's view to theoretical reasons, we'd have to conclude that this person doesn't have a reason to believe what the available evidence supports. But that's implausible. It's better to say instead that she does have a reason to believe what the available evidence supports, but she is insensitive to such reasons. (After all, if there were no such reasons, why think there's anything *problematic* about her being dogmatic, brainwashed, or prone to wishful thinking? These conditions seem problematic largely because they render a person insensitive to her reasons for belief.) This yields pressure to have an analogous account of practical reasons.

A second strategy would be to adopt Manne's own "practice-based" approach to normativity, but consider practices besides that of reasoning with others in which reasons make an appearance. On Manne's practice-based approach, we start our theorizing by looking at our normative practices and behaviors, and then define our normative concepts in terms of the roles played within those practices:

> 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. … 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.[45]

Manne focuses specifically on our practice of *reasoning with* others, and this leads her to think that reasons apply to a person only if there's a point in offering them to that person. But we could instead start with another practice: our practice of *reasoning about* the actions of (actual or hypothetical) others. For instance, I might have a discussion with my friend Alice, or think to myself, about whether our mutual friend Bernie should attend UCLA. In the course of this discussion, Alice and I might list out some pros and cons of Bernie's attending UCLA.[46] (We could even imagine an idealized version of this practice in which there are no limitations as far as information, time, and resources go, and so no need to restrict ourselves only to the weighty pros and cons.) In engaging in this practice of listing out the pros and cons, it seems natural to think that what we're doing is listing out the *reasons* for and against Bernie's attending UCLA. (These pros and cons, after all, seem to have all the features of reasons: they contribute to some all-things-considered verdict about what he should do; they can be outweighed; they can be undercut; and so forth.) Unlike the practice of *reasoning with* Bernie, in *reasoning about* Bernie, we are not constrained by whether there would be a point in presenting those considerations to Bernie. In *this* practice, we need not consider how well the considerations would resonate with Bernie before including them on the "pro" or "con" list.

If we consider an idealized *reasoning-about* practice with respect to the question of whether Bernie should attend UCLA, we'd get one set of pros and cons. But if we instead considered an idealized *reasoning-with* practice with respect to the question of whether Bernie should attend UCLA, we would get a subset of those pros and cons, since we'd exclude those pros and cons, if any, that wouldn't resonate with Bernie. But shouldn't we reflect on idealized versions of these two practices *together*? Since we're interested in understanding what reasons are, and both practices appear to employ the concept of a

reason, we should look to *both* practices to understand what reasons are. The problem with Manne's strategy is that she looks at just one of these practices, the reasoning-with practice, and concludes that some fact can be a reason only if it would achieve uptake in idealized conversation. That's like looking just at National League baseball, and concluding that one can be pitching in a baseball game only if one has a spot in the batting order.

In summary, there are a couple of strategies available to the externalist who replies to Manne by insisting, roughly, that the truth of reasons claims is one thing, and their rhetorical effectiveness another. The externalist can point out that Manne's view would be implausible when applied to theoretical reasons. And the externalist can argue that even within Manne's "practice-based" framework, a wider view of which practices are relevant wouldn't license the internalist restrictions Manne favors.

There are other ways in which the distinction between the *truth* of reasons claims and their *rhetorical effectiveness* might be relevant for debates over internalism. As we've seen, the externalist will claim that we cannot infer from a reasons claim being *rhetorically ineffective* that the reasons claim is false. Even if the nasty husband is unmoved by any attempt to persuade him that his wife's needs and interests constitute a reason for him to be nice to his wife, it could still be true that this is a reason for him to be nice to his wife. Likewise, the externalist will claim that we cannot infer from a reasons claim being *rhetorically unnecessary* that the reasons claim is false. Suppose a very considerate husband *doesn't need* to be persuaded to be nice to his wife. We can suppose that even deliberating about the pros and cons of being nice to his wife strikes him as distasteful. But it could nonetheless be *true* that his wife's needs and interests constitute a reason for him to be nice to his wife. He just doesn't need this explained to him.

The latter point is relevant to Alan Goldman's defense of internalism. Goldman aims to defend internalism against an extensional objection: internalism is unable to recognize a class of reasons that intuitively exists – namely, reasons for intrinsic desires. Suppose I desire that my children do well, and not because their doing well is a means to something else I desire. Internalists can allow that this desire provides me with a reason to do various things to promote the welfare of my children, such as enrolling them in good schools, which might in turn provide me with a reason to live in a good school district, and so forth. But internalists cannot allow that there are reasons for me to desire that my children do well. As Derek Parfit puts the objection, internalists must think that "any chain of reasons must end with some desire that we have no reason to have."[47]

One avenue of reply would be to develop a Humean view that isn't committed to holding that any chain of reasons must end with some desire that we have no reason to have. (Mark Schroeder's Hypotheticalism, for instance, is consistent with someone's having reasons for every desire she has.[48]) Goldman, however, replies by arguing that our deepest concerns "need no reasons ... and for some of them, reasons are out of place."[49] He provides some examples:

> I need no reason to love my wife or children, to care about my own health or welfare, to want to have a successful career. In some of these cases I could have reasons to have these concerns. I might have a reason to be concerned about my health in order to preserve myself for writing my philosophical masterwork. But I need no reason to be rationally concerned about my health. In other cases reasons seem more strongly out of place. Citing reasons to love my children would be worse than odd.[50]

Goldman is right that we don't "need" reasons to love our children since we don't take the question of whether to love them as open for deliberation. Such reasons are rhetorically unnecessary. But, to respond to the extensional objection, what Goldman needs to show is that, despite our intuitions to the contrary, *there are no* reasons for one's (instrinsic) love of one's children. And, as we've seen, it doesn't follow from some consideration's being rhetorically unnecessary that it's not a reason.

Additionally, if it's inappropriate, and perhaps even incompatible with virtue, to go through the pros and cons of loving your children, we are not forced to conclude that those reasons claims that are inappropriate to consider are *false*. It just means that there are some truths that sometimes shouldn't be on one's mind. We could put the point another way, borrowing Bernard Williams's memorable phrase: the person who fails by having "one thought too many" could nonetheless have a thought that's true.[51]

3. GOOD DELIBERATION

Another way to argue for internalism is to start with a conceptual connection between an agent's reasons and *good deliberation*: roughly, one's reasons must motivate one to act when one is deliberating well. Hille Paakkunainen (forthcoming) formulates this idea as the

> *Deliberative Constraint*: The fact that p is a (decisive) normative reason for A to ϕ only if A would ϕ because p if A deliberated well.[52]

If we then understand "deliberating well" in roughly the minimal way Williams himself understood it, in which the sound deliberative route must *proceed from* the agent's existing motivations, we arrive at an internalist view of reasons.[53] Williams held that deliberating well involves, among other things, correction for mistaken beliefs. In his famous example

of someone who wants a drink of gin and falsely believes that the glass in front of him contains gin when it actually contains gasoline, we can say that if he "deliberated well" – that is, if he had true beliefs about what's in the glass – he wouldn't be motivated to drink the contents of the glass, and so has no reason to drink the contents of the glass. While correct deliberation wouldn't give any weight to desires based on false beliefs (like the desire to drink the contents of the glass), there is no similar correction for imprudent or immoral desires.

One way to resist the internalist conclusion is to hold onto the Deliberative Constraint and adopt a more expansive conception of good deliberation. John McDowell, for instance, holds that to deliberate well, one must also have the motivations that a virtuous person has.[54] Combined with *this* conception of good deliberation, the Deliberative Constraint wouldn't pose any threat to the existence of categorical reasons: the nasty husband *does* have a reason to be nice to his wife since he would be motivated to be nice if he deliberated well in *this* sense. But this strategy puts all the weight of the argument on the defense of the more expansive conception of good deliberation.

Another way to resist the internalist conclusion is to deny the Deliberative Constraint, which has come in for criticism lately, even from some internalists. Markovits (2014), who defends internalism on other grounds, argues that the Deliberative Constraint should be rejected largely because of "conditional fallacy" concerns.[55] The worry here is that if we construct our account of reasons simply by looking at how someone would be motivated if they were deliberating well, we would miss out on those reasons explained by their failure to deliberate well. To take a standard example, if I were deliberating perfectly well, I wouldn't be motivated to improve my deliberations. But this may be something that I, with all my imperfections, have reason to do.[56]

Here's Markovits's presentation of a case, originally due to Robert Johnson, that appears to be a counterexample to the Deliberative Constraint:

> Let's say I become convinced I am James Bond. The fact that I am suffering from such a delusion may give me excellent reason to see a psychiatrist for treatment. But it cannot motivate me to see the psychiatrist. For if this fact could motivate me to seek help, I would no longer be convinced I was James Bond. Someone who firmly believes he is James Bond cannot be motivated to seek a psychiatrist by the fact that his belief is a delusion.[57]

The counterexample isn't airtight, since, as others have observed, one need not lose the belief that one is James Bond in becoming aware of the delusion.[58] After all, if we allow for the possibility of irrational inconsistency in belief – including even, as some claim, the possibility of believing that P and believing that not P at the same time – then it's not clear why we must think it's impossible for me to believe I'm James Bond and to believe I'm deluded. If that's possible, then I could deliberate from the awareness of my delusion to the conclusion to seek help. However, there are structurally similar examples that will work just as well. Here's one from Mark Schroeder:

> Nate loves successful surprise parties thrown in his honor, but can't stand unsuccessful surprise parties. If there is an unsuspected surprise party waiting for Nate in the living room, then plausibly there is a reason for Nate to go into the living room. There is certainly something that God would put in the "pros" column in listing the pros and cons of Nate's going into the living room. But it is simply impossible to motivate Nate to go into the living room for this reason – for as soon as you tell him about it, it will go away.[59]

This case appears to be a counterexample to the Deliberative Constraint: the fact that there's a surprise party in the living room is a reason for Nate to go into the living room, but it's not the case that Nate would go into the living room because there's a surprise party if he deliberated well.

However, "James" and Nate have not settled the debate. One strategy open to defenders of the Deliberative Constraint is to deny that "James" has a reason to seek help and that Nate has a reason to go into the living room. In support of this, we might suggest that there are closely related evaluative or normative claims which are true of them: perhaps it would be *good* if "James" sought help, and if Nate went into the living room, without it being the case that either have reasons to do so.[60] And perhaps, as Kieran Setiya suggests in discussing a related example, the surprise party is a reason for Nate *to be glad* if he goes into the living room, and *dismayed* if he doesn't, but not a reason for him *to go* into the living room.[61] This strategy holds that "James" and Nate aren't counterexamples to the Deliberative Constraint, and explains the temptation to think they are: we've mistakenly thought the relevant agents had reasons for action, when we should have instead applied some closely related evaluative or normative concept (what would be good for them, what they have reason to be glad or dismayed about, or perhaps something else).

Additionally, Paakkunainen offers an argument for the conclusion that "James" and Nate don't have the relevant reasons for action. She first observes that in such cases, "it would be a sheer accident if I happened to find myself doing what I supposedly have decisive reason to do." For instance, it would be a sheer accident if Nate ended up at the surprise party, and if "James" ended up in the psychiatrist's office. Moreover, *no one* in such circumstances could end up doing what they supposedly have decisive reason to do

except by sheer accident – no matter how well-informed they were nor how well they deliberated. She then claims, "But the idea that no-one could come to do what they ought to do except by a happy accident, no matter how wise and well-informed they were and no matter how well they deliberated, is hard to believe."[62]

Given these challenges, critics of the Deliberative Constraint need to say more about why "James" and Nate are genuine counterexamples. I'll take up that task here. (I won't be able to provide anything like a comprehensive assessment of the Deliberative Constraint in this limited space. So, I'll limit myself to giving reasons to think these alleged counterexamples do indeed pose a serious difficulty for it.) I'll focus on Nate, but everything could be said of "James" as well.

First, let's consider why we should think there's *a reason* for Nate *to go* into the living room, as opposed to merely a reason to be glad if he does and dismayed if he doesn't, or merely good for him to go. The answer, I think, is that the concept of a reason links up to the concept of *ought* in ways that these other normative concepts do not. It's commonly accepted that one ought to do what there is *most reason* to do. On this view, we "weigh up" the relevant reasons for action to reach some all-things-considered verdict about what one ought to do. So, a change in reasons for and against ϕ-ing can impact whether one ought to ϕ. But these other normative concepts do not share this same conceptual connection with what one ought to do. Now, intuitively, Nate ought to go into the living room. (We're here using the same objective sense of "ought" that's used when we say that the agent in Williams's example ought not drink the contents of the glass – the sense of "ought" that would be used by an informed advisor.) That means that there is most reason for him to go into the living room. But if we now say that, because of

the Deliberative Constraint, the surprise party isn't a reason for Nate to go into the living room, and (we'll assume) there are no other reasons for him to go into the living room, then we're forced to conclude that there isn't most reason for him to go there, and hence it's not the case that he ought to go there. We've reached a counterintuitive conclusion. To avoid that conclusion, it won't be enough to say that there's a reason for Nate to be glad if he goes and dismayed if he doesn't, or to say that it would be good if he goes. Rather, we need to say that there's a reason for him to go into the living room.

Second, the explanation that the defender of the Deliberative Constraint would have to give for why it's not the case that Nate ought to go into the living room doesn't seem to track anything of normative significance. To see this, it would help to consider a related case. Kate, like Nate, loves successful surprise parties thrown in her honor. But, unlike Nate, she doesn't mind unsuccessful surprise parties; if the surprise were spoiled, she'd still be motivated to go. But let's suppose that the surprise isn't spoiled, Kate goes into the living room, and she enjoys the successful surprise. Presumably, both sides to this dispute would agree that Kate acted as she ought to have acted (again, in the objective sense of "ought") by going into the living room, since the Deliberative Constraint doesn't rule out any of Kate's reasons. Now suppose Nate also goes into the living room and enjoys the successful surprise. Defenders of the Deliberative Constraint, who are committed to thinking that it's not the case that Nate ought to have gone, would have to offer this explanation for the normative difference between Nate and Kate: had the surprise been spoiled, Nate wouldn't have been motivated to go. But how is this fact normatively significant? The surprise *wasn't* spoiled, and it's how things went, not how they could have went, that's normatively significant for what Nate (objectively) ought to have done.[63] In short, the explanation that defenders of the Deliberative Constraint

would have to give to distinguish Nate from Kate doesn't track anything of normative significance.[64]

Let's return to Paakkunainen's argument. I think she's right that Nate could end up doing what he ought to do only by sheer accident. And she might also be right that this is a cost that opponents of the Deliberative Constraint might have to bear. But how significant a cost is it? When we're using the objective sense of "ought," we're already quite comfortable with the possibility of people doing what they ought to do by sheer accident. In Williams's gin and gasoline example, the agent might get distracted and refuse the drink, thereby doing what he ought to do, by sheer accident. And if we make a slight change – an undetectable poison, instead of gasoline – it would be only through sheer accident that he could do what he ought to do and not drink it. In this case, it's still true that a fully informed version of the agent could deliberate his way to not drinking it, and that distinguishes this case from Nate's case. But, *absent the miraculous provision of information*, he'll do what he ought to do only by sheer accident. And if we're already comfortable with this thought, then it's not too significant a cost, if any, to also allow that sometimes it's impossible for one to do what one ought to do except by sheer accident.

In summary, the Deliberative Constraint, which might serve as a premise in an argument for internalism, appears subject to counterexamples. Although there is controversy about these counterexamples, I've suggested some reasons for thinking they do pose serious problems for the Deliberative Constraint.

4. CONCLUSION

In this paper, I've looked at some recent interesting work on internal and external reasons. Specifically, I've considered arguments for internalism based on i) the analogy

with theoretical reasons, ii) the relationship between reasons and our practice of reasoning with others, and iii) the relationship between reasons and good deliberation. These important developments are more evidence that there are many interesting possibilities for defending internalism besides the argument Williams offered. However, I've presented reasons to be skeptical of each of these developments; I don't think they provide us with sufficient reason to believe internalism is true.

University of Nebraska - Lincoln

REFERENCES

Brady, Michael. 2000. "How to Understand Internalism," *Philosophical Quarterly*, vol. 50, pp. 91-97.

Brunero, John. 2008. "McDowell on External Reasons," *European Journal of Philosophy* vol.16, pp. 22-42.

Clark, Philip. 2001. "Velleman's Autonomism," *Ethics*, vol. 106, pp. 580-593.

Cohon, Rachel. 1986. "Are External Reasons Impossible?," *Ethics*, vol. 96, pp. 545-556.

Darwall, Stephen. 1992. "Internalism and Agency," *Philosophical Perspectives (Ethics),* vol. 6, pp. 155-174.

Finlay, Stephen. 2009, "The Obscurity of Internal Reasons," *Philosophers' Imprint,* vol. 9, pp. 1-22.

Finlay, Stephen and Mark Schroeder. 2012. "Reasons for Action: Internal vs. External," in Edward N. Zalta (ed.) *Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), URL = http://plato.stanford.edu/archives/win2012/entries/reasons-internal-external/.

Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives," *Philosophical Review*, vol. 81, pp. 305-316.

------. 2001. *Natural Goodness*. (Oxford, UK: Oxford University Press).

Gert, Joshua. 2002. "Avoiding the Conditional Fallacy," *Philosophical Quarterly,* vol. 52, pp. 88-95.

Gibbons, John. 2013. *The Norm of Belief* (Oxford, UK: Oxford University Press).

Goldman, Alan. 2009. *Reasons from Within: Desires and Values* (Oxford, UK: Oxford University Press).

Graham, Peter A. 2010. "A Defense of Objectivism about Moral Obligation," *Ethics,* vol. 121, pp. 88-115.

Grice, Paul. 1975. "Logic and Conversation," in P. Cole and J.L. Morgan (eds.) *Syntax and Semantics 3: Speech Acts* (New York, NY: Academic Press), pp. 41-58.

Harman, Gilbert. 1975. "Moral Relativism Defended," *Philosophical Review*, vol. 84, pp. 3-22.

Hooker, Brad. 1987. "Williams' Argument Against External Reasons," *Analysis,* vol. 47, pp. 42-44.

Hyun, Alexander. 2015. "Internalism, Ideal Advisors and the Conditional Fallacy," *Journal of Ethics and Social Philosophy*, www.jesp.org, Discussion Note, August 2015.

Johnson, Robert. 1999. "Internal Reasons and the Conditional Fallacy," *Philosophical Quarterly,* vol. 49, pp. 53-71.

------. 2003. "Internal Reasons: Reply to Brady, van Roojen, and Gert," *Philosophical Quarterly,* vol. 53, pp. 573-580.

Joyce, Richard. 2001. *The Myth of Morality* (Cambridge, UK: Cambridge University Press).

Kelly, Thomas. 2003. "Epistemic Rationality as Instrumental Rationality: A Critique," *Philosophy and Phenomenological Research,* vol. 66, pp. 612-640.

Korsgaard, Christine. 1986. "Skepticism about Practical Reason," *Journal of Philosophy,* vol. 83, pp. 5-25.

Leech, Geoffrey. 1983. *Principles of Pragmatics* (London, UK: Longman).

Mackie, John. 1977. *Ethics: Inventing Right and Wrong* (Harmondsworth, UK: Penguin Books).

Manne, Kate. 2014. "Internalism about Reasons: Sad but True?," *Philosophical Studies*, vol. 167, pp. 89-117.

Markovits, Julia. 2014. *Moral Reason* (Oxford, UK: Oxford University Press).

McDowell, John. 1995. "Might There be External Reasons?," in J.E.J. Altham and Ross Harrison (eds.) *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge, UK: Cambridge University Press), pp. 68-85.

Millgram, Elijah. 1996. "Williams' Argument Against External Reasons," *Noûs*, vol. 30, pp.197-220.

Paakkunainen, Hille. Forthcoming. "Internalism and Externalism about Reasons," in Daniel Star (ed.) *Oxford Handbook on Reasons and Normativity* (Oxford, UK: Oxford University Press).

Parfit, Derek. 1997. "Reasons and Motivation," *Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 71, pp. 99-146.

Scanlon, T.M. 1998. *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Schroeder, Mark. 2007. *Slaves of the Passions* (New York, NY: Oxford University Press).

Sepielli, Andrew. Forthcoming. "Subjective and Objective Reasons" in Daniel Star (ed.) *Oxford Handbook on Reasons and Normativity* (Oxford, UK: Oxford University Press).

Setiya, Kieran. 2009. "Reply to Bratman and Smith," *Analysis,* vol. 69, pp. 531-540.

------. 2012. "Introduction: Internal Reasons," in Kieran Setiya and Hille Paakkunainen (eds.) *Internal Reasons: Contemporary Readings* (Cambridge, MA: MIT Press), pp. 1-34.

Smith, Michael. 1994. *The Moral Problem* (Oxford, UK: Blackwell Publishers).

------. 1995. "Internal Reasons," *Philosophy and Phenomenological Research,* vol. 60, pp. 109-131.

------. 2009. "Reasons With Rationalism After All," *Analysis*, vol. 69, pp. 521-530.

Skorupski, John. 2007. "Internal Reasons and the Scope of Blame," in Alan Thomas (ed.) *Bernard Williams* (Cambridge, UK: Cambridge University Press), pp. 73-103.

Thomas, Alan. 2006. *Value and Context* (Oxford, UK: Oxford University Press).

van Roojen, Mark. 2000. "Motivational Internalism: A Somewhat Less Idealized Account," *Philosophical Quarterly,* vol. 50, pp. 233-241.

Velleman, David. 1996. "The Possibility of Practical Reason," *Ethics,* vol. 106, pp. 694-726.

------. 2000. "Introduction" in *The Possibility of Practical Reason* (Oxford, UK: Oxford University Press).

Way, Jonathan. Forthcoming. "Reasons as Premises of Good Reasoning," *Pacific Philosophical Quarterly*.

Wiland, Eric. 2012. *Reasons* (London, UK: Continuum).

Williams, Bernard. 1981a. "Internal and External Reasons," reprinted in *Moral Luck* (Cambridge, UK: Cambridge University Press), pp. 101-113.

------. 1981b. "Persons, Character, and Morality" reprinted in *Moral Luck* (Cambridge, UK: Cambridge University Press), pp. 1-19.

------. 1995a. "Internal Reasons and the Obscurity of Blame," reprinted in *Making Sense of Humanity* (Cambridge, UK: Cambridge University Press), pp. 35-45.

------. 1995b. "Replies," in J.E.J. Altham and Ross Harrison (eds.), *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge, UK: Cambridge University Press), pp. 185-224.

------. 2001. "Postscript: Some Further Notes on Internal and External Reasons," in Elijah Millgram (ed.) *Varieties of Practical Reasoning* (Cambridge, MA: MIT Press), pp. 91-97.

NOTES

[1] Thanks to the graduate students in my Fall 2014 seminar on internal reasons at UNL, and to Mark van Roojen and Joe Mendola, for attending that seminar and for many useful conversations. And thanks to Mark, and to John Greco, for helpful comments on a previous draft of this paper.

[2] Williams (1981a), pp. 102, 105. See also Williams (1995a, 1995b, 2001). I'll here follow Williams (1981a), p. 105, and use the term "desires" to refer to all of these elements of the subjective motivational set, including desires in the ordinary sense.

[3] Williams (1981a), p. 101, and (1995a), p. 35.

[4] Williams (1995a), p. 35. Williams's internalism is what Stephen Darwall (1992), p. 155, helpfully refers to as "existence internalism," since it concerns a necessary condition for the existence of a reason, in contrast to "judgment internalism," which is a separate thesis about the necessary connection between normative judgment and motivation, and which will not be our concern here, except to briefly note the role it plays in an argument for existence internalism discussed at the start of § 2. In this paper, I'll stay close to Williams's usage of "internalism" and "externalism." Other terminological choices are possible. For instance, Finlay and Schroeder (2012), § 1.1, present a schema for internalism that would include Williams's view, but also extend to many others: "Schematic Internalism: Every reason for action must bear relation *R* to motivational fact *M*." This schema would, as Finlay and Schroeder note in § 1.2, have the unusual result that some views thought to be classic *externalist* views, like McDowell (1995), count as internalist. Nothing of substance will hinge on our choice of terminology here.

[5] Williams (1981a), p. 104. Since Williams's conception of sound deliberation involves, among other things, correcting for mistaken beliefs (see p. 102), it's clear that he's giving us an account of *objective*, not subjective, reasons. For a discussion of this distinction and issues related to it, see Sepielli (forthcoming). Unless otherwise specified, we'll be discussing objective reasons, and the related notion of the objective "ought," in this paper.

[6] Williams (1995a), p. 39.

[7] I'll here use the terminology of "categorical reasons" instead of the equivalent "external reasons" since it better brings to mind this connection with our ordinary understanding of morality.

[8] Mackie (1977), Ch. 1, especially § 6; Joyce (2001), p. 42.

[9] Harman (1975), p. 9.

[10] Foot (1972). In rejecting the conclusions of that paper in her later work, Foot (2001), p. 61, writes: "I now wonder why, given the obvious indigestibility of the idea that morality is indeed a system of hypothetical imperatives, I should have accepted it even for a short while. What seemed to force it on me was the sheer difficulty of showing a practical rationality that was independent of desire or interest."

[11] See, for instance, Smith (1995), Schroeder (2007), and Markovits (2014).

[12] Williams (2001), p. 93, tells us that his argument against external reasons appeals to the claim that "…if it is said, in the normative mode, that A has a reason to ϕ, the speaker must envisage the possibility of A's ϕ-ing for that reason, in which case the reason will figure in the explanation of what A does." This claim is false, since it is vulnerable to the same counterexamples to the Deliberative Constraint that I discuss in § 3 of this paper, and hence Williams's argument against external reasons is unsound.

[13] Some important discussions of Williams's argument can be found in Cohon 1986; Finlay 2009; Hooker 1987; McDowell 1995; Millgram 1996; Paakkunainen, forthcoming; Scanlon 1998, pp. 363-373; Setiya 2012; Skorupski 2007; Thomas 2006, pp. 67-100; and Wiland 2012, pp. 31-37. See also Finlay and Schroeder 2012, which includes an extensive bibliography on internal and external reasons.

[14] A good deal of that renewed interest is due to Mark Schroeder's (2007) much-discussed Humean theory of reasons, Hypotheticalism, which we won't discuss in this paper.

[15] See Finlay and Schroeder (2012), §2.2, and Kelly (2003).

[16] This example is from Kelly (2003), p. 626.

[17] Goldman (2009), p. 72.

[18] Goldman (2009), p. 68. See also Velleman (1996).

[19] Velleman (2000), p. 19. See also Kelly (2003), pp. 631-632.

[20] Goldman (2009), p. 71.

[21] Goldman (2009), p. 72.

[22] Goldman (2009), p. 72. David Velleman (1996) has advanced a structurally similar argument that identifies the constitutive aim of action as autonomy. That constitutive aim wouldn't support the internalist conclusion which Goldman wants to draw, and is rejected by Goldman for independent reasons. See Goldman (2009), pp. 69-70. See also the important criticisms of Velleman's proposal in Clark (2001).

[23] Goldman (2009), p. 73.

[24] Velleman (2000), p. 16.

[25] Markovits (2014), p. 60.

[26] Markovits (2014), p. 63.

[27] This is a variation on an example Markovits uses, on p. 7, in a different context.

[28] See Gibbons (2013), p. 179, for another example with a similar structure.

[29] Perhaps there's some flexibility in the notion of "procedural rationality" such that being procedurally rational involves one's having the beliefs one should have. If so, we can say that a procedurally rational physician would believe the patient is allergic, and so our physician *can* reach the conclusion that the drug will make the patient worse off through procedurally rational deliberation from her current beliefs. However, I doubt that Markovits would allow for this to count as *procedural*, as opposed to *substantive*, rationality. (See how she draws the distinction on p. 110.) Moreover, this expanded conception of procedural rationality would help the *externalist's* case, since the externalist could then say that the practical analogue to having the beliefs one should have is having the motivations one should have. And if we build into our conception of practical rationality the having of the motivations one should have, it's easy to allow for the existence of categorical reasons. (Indeed, John McDowell (1995) defends an Aristotelian externalist view precisely along these lines.)

[30] When it comes to the physician's *practical* reasons, many philosophers would claim that although she had no *objective* reason to prescribe the drug (since it was actually harmful), she had strong *subjective* reasons to do so (since, given her evidence, that option was expected to produce the best results). Very roughly, objective reasons are provided by the actual outcomes, while subjective reasons are provided by expected outcomes, given the available evidence, or the subject's beliefs, or something similar. Whereas Williams's internalism was clearly a view about objective reasons – see his famous gin/petrol example in (1981a), p. 102 – Markovits formulates her internalism as a view about

subjective reasons, since these are the reasons which she takes to be connected to moral obligations. As she puts it, "it's plausible that we are morally obligated to do only what we have sufficient evidence to believe it would be best to do, not what it would (in fact) be best to do" (p. 7). (This is a controversial view. See Graham (2010) for a defense of the opposing view.) The view that moral obligations track subjective, and not objective, reasons would also share the same source of intuitive support mentioned in the text: the physician *isn't an appropriate target for blame or criticism*, since she proceeded impeccably from the evidence available to her in prescribing the drug.

[31] Markovits thinks that internalism and the traditional understanding of morality are compatible, and, in the second half of her book, she presents an ingenious variation on Kant's argument for the Formula of Humanity in support of this view. So, she would hold that there is actually a way, via procedurally rational deliberation from the uncaring physician's motivations, to arrive at the conclusion not to prescribe the drug. But this doesn't affect the main point here: *if there weren't* a procedurally rational link between the uncaring physician's motivations and the conclusion not to prescribe the drug, *she wouldn't be let off the hook*. And that's enough to establish a difference with the theoretical case.

[32] Although I won't discuss judgment internalism here, this version of it strikes me as implausibly strong. When it comes to reasons to ϕ that are severely outweighed, it seems I can recognize the existence of these reasons without having any motivation whatsoever to ϕ. (For an example, see Mark Schroeder's case of Joel in Schroeder (2007), pp. 166-167.)

[33] See Manne (2014), p. 109, for the detailed version of the argument.

[34] Manne (2014), p. 109.

[35] Manne (2014), p. 97.

[36] Manne (2014), p. 97.

[37] Manne (2014), p. 91.

[38] I'll leave out this qualification in what follows. Just assume that when I speak of the ideal advisor reasoning with the advisee, I mean that she is reasoning with her in an ideal way.

[39] Manne (2014), p. 95.

[40] Manne (2014), p. 95.

[41] Manne (2014), p. 102.

[42] See Manne (2014), p. 106.

[43] See Grice (1975), pp. 45-47, and Leech (1983), esp. chapters 4-6.

[44] Manne (2014), p. 112. Manne's claim here seems vulnerable to the alleged counterexamples to the Deliberative Constraint discussed in § 3 of this paper. Additionally, as Alexander Hyun (2015) argues, Manne's version of internalism seems subject to versions of the "conditional fallacy" objection.

[45] Manne (2014), p. 94.

[46] Some of these considerations may be moral ones and some may not. So, this practice is distinct from the practice, discussed by Manne on p. 101, of constructing a "moral report card" for the agent, though no doubt pros and cons might be listed out in that practice as well.

[47] Parfit (1997), p. 127-128.

[48] Schroeder (2007), p. 186.

[49] Goldman (2009), p. 129.
[50] Goldman (2009), pp. 128-129.
[51] Williams (1981b), p. 18.
[52] Paakkunainen (forthcoming), msp. 18. Although this formulation concerns decisive reasons, perhaps something similar could be formulated for outweighed reasons. See Way (forthcoming) for an important discussion relevant to this task.
[53] Williams thinks deliberation must proceed from the agent's subjective motivational set. See (1981a), pp. 104, 109.
[54] McDowell (1995), p. 78. See Brunero (2008) for further discussion of McDowell's view. One might read Korsgaard (1986) as alerting us to the possibility of a more ambitious, Kantian conception of good deliberation.
[55] See Johnson (1999) and Johnson (2003), which is a reply to Brady (2000), van Roojen (2000), and Gert (2002).
[56] Williams (1995), p. 190, presents this kind of worry against McDowell's externalism. See also Michael Smith's critique of the "example model" in Smith (1995), pp. 110-112.
[57] Markovits (2014), p. 41. See Johnson (1999), pp. 66-67, and Johnson (2003). Smith (2009), p. 523, also makes this point with the example of someone who believes he is Jesus.
[58] See Paakkunainen (forthcoming), msp. 29 and Setiya (2009), pp. 538.
[59] Schroeder (2007), p. 165.
[60] Paakkunainen (forthcoming), msp. 30.
[61] Setiya (2009), pp. 538.
[62] Paakkunainen (forthcoming), msp. 30.
[63] Suppose that instead of surprise parties, Nate enjoys parties with loud music. The fact that there's loud music at the party provides him with a reason to go. It doesn't matter that it could have been irritatingly quiet.
[64] There's another strategy available to defenders of the Deliberative Constraint: to hold that Nate *does* have a reason to go into the living room, but it's not *that there's a surprise party*, but something else, perhaps *that doing so would be fun*. For this to be a reason, the Deliberative Constraint requires that Nate would go into the living room because doing so would be fun, if he deliberated well. We could hold fixed the ground of Nate's reason (his ignorance of the surprise), and then say that if Nate deliberated well, where this involves his knowing that going into the living room would be fun, he would go into the living room for this reason. (This applies a strategy from van Roojen 2002.) *This* reason thus satisfies the Deliberative Constraint. And so Nate, like Kate, ought to have gone into the living room. There are a couple of difficulties here. First, opponents of the Deliberative Constraint could respond by altering contingent facts about Nate's circumstances and psychology so that he *wouldn't* go into the living room because doing so would be fun, if he deliberated well. Perhaps his living room is so decrepit and dull that the only way there's any fun to be had in there is if there's a surprise party. Knowing this, Nate should reason that since going into the living room would be fun, there must be surprise party waiting for him, and so (the surprise now spoiled) going in would be no fun after all. Here, good deliberation would *not* lead Nate to go into the living room for this reason, and so the Deliberative Constraint would rule this reason out as well. Second, even supposing the Deliberative Constraint *doesn't* rule out this reason, we'd still have the problem of *normative insignificance*. Let's suppose unsuspecting Nate goes into the living

room and encounters a wonderful surprise party, with good friends, music, etc. In going into the living room, Nate did what he (objectively) ought to have done. Now, it's easy to see how some *changes in how things actually went* could change the verdict about what Nate ought to have done. If we alter features of the party – make the attendees his enemies instead of his friends – or make it such that he promised not to move, we could alter the verdict about what he ought to have done. These features have normative significance. But it harder to see how changes, not in how things went, but in *how they would have gone had Nate known that going into the living room would be fun* – specifically, whether his awareness would allow for him to enter the living room for this reason or not – should matter for what Nate ought to have done. Unlike facts about who or what is at the party, or what Nate has promised, facts about the deliberative avenues that would be open to Nate, were he, contrary to fact, aware that going into the living room would be fun, seem *normatively insignificant*. Yet if the Deliberative Constraint is right, what Nate ought to do depends on such facts.