

Reason Internalism

ALAN H. GOLDMAN

College of William & Mary

This paper defends strong internalism about reasons, the view that reasons must relate to pre-existing motivational states, from several kinds of counterexamples, supposed desire independent reasons, that have been proposed. A central distinction drawn is that between there being a reason and an agent's having a reason. For an agent to have an F reason, she must be F-minded. Reasons, as what motivate us, are states of affairs and not themselves desires or motivational states, but they must connect to existing motivational states. It has been claimed that rationality itself requires us to recognize certain reasons independent of our desires, that we acquire new desires by learning what is valuable, by acquiring desire-independent reasons to pursue certain values. It is claimed also that prudential and moral reasons are desire independent. By offering an account of rationality as coherence, by appealing to broader concerns as opposed to specific desires, and by appealing to the distinction noted above, the paper exposes weaknesses in recent arguments for desire independent reasons by Millgram, Smith, Korsgaard, and Searle. The reasons they propose can be interpreted as internal (not desire independent) or dismissed as nonexistent.

I. Groundwork

All internalists about practical reasons hold that such reasons motivate rational agents. There is a stronger claim in the doctrine as it will be understood here, namely, that a necessary condition for a consideration's being a reason is that it relates to an existing concern or desire of the agent. It is because agents have certain concerns or desires that they have reasons; their motivational states limit what reasons they can have. A reason is not a reason intrinsically: in itself it cannot demand on pain of irrationality that an agent be motivated by it.¹ The internalist in this strong sense can allow that an agent has a reason even though she is not motivated by it at a particular

¹ John Robertson, "Internalism, Practical Reason, and Motivation," in E. Millgram, ed., *Varieties of Practical Reasoning* (Cambridge, MA: MIT Press, 2001) draws a similar distinction with very different labels. He calls "attenuated internalism" what I am calling internalism. What he calls "robust internalism," the position that holds that reasons in themselves demand motivation, I am refusing to call internalism. My usage is more in line with what Bernard Williams had in mind when he first defended internalism in "Internal and External Reasons," in R. Harrison, ed., *Rational Action* (Cambridge: Cambridge University Press, 1979).

time. She may be in an irrational frame of mind at that time, too depressed, for example, to care about any of her usual concerns. But such concerns there must be in relation to which the reason is a reason, before the lack of motivation can count as irrational. On the other side, opponents of internalism in this strong sense can allow us to posit desires whenever we are motivated to act on reasons, but they will typically see these desires as themselves motivated by the reasons in themselves, as are the actions that result.²

The underlying rationale for internalism in this sense appeals to the nature and origin of value. It is the recognition that values exist only for and because of valuing agents, agents motivated to seek or to avoid certain states of affairs.³ Values are relations between objects or states of affairs and subjects who value them. This is, of course, a controversial view. Many philosophers believe in objective values, things that agents ought to value (potential motives) whether they do or not. But that values in objects depend on actual valuing by agents is the underlying premise of internalists, for whom the demand to value various objects follows only from the demand for coherence with existing concerns and motivations. If things have value only in relation to the concerns of valuing agents, then reasons are such only relative to these same concerns. For reasons, as considerations that count for or against particular actions, are perceived positive or negative values, reflecting ends sought or shunned. Reasons are what motivate us as rational agents, but they motivate us by reflecting our values, ends, desires or concerns. As considerations that motivate us, they are naturally seen as the contents of our desires, although we are typically aware of these contents, good and bad states of affairs, and not of the desires or motivational states themselves, except in the absence of some strongly desired state of affairs. Given this fundamental claim about values as premise, internalism seems to follow.⁴ But the question for this paper is whether it can be maintained in light of apparent counterexamples, reasons that seem to exist, or that we seem to have, independent of our motivations. Strong reason internalism is on the defense at present, and I intend to show what defense it can offer.

As a first step toward explaining away these purported counterexamples, the following distinction between definitions will be crucial: (1) *There is an F (moral, prudential, aesthetic...) reason R to do act A = If (and only if) a subject S is F-minded, then S, if rational, would be motivated by awareness of R to A;* (2) *S has an F reason R to A = S is F-minded and, if rational, would be motivated by awareness of R to A.* The first clause on the right side of the second definition may be redundant, since S's motivation implies his being

² See, for example, Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970); Joseph Raz, *Engaging Reason* (Oxford: Oxford University Press, 1999).

³ This broad definition allows that most sentient creatures are valuing subjects.

⁴ But Christine Korsgaard, to be discussed below, accepts the premise but argues that we must be morally concerned whatever our antecedent motivations.

F-minded or concerned. The first clause on the right side of the first definition reflects the internalist thesis that F reasons require F concerns on the part of the subject. If a particular tennis racket is on sale at the Sports Authority store, then there is a reason for me to go there if and only if I am concerned about my tennis game and about saving money while I acquire the means to improve it. The second definition allows that subjects can have reasons of which they are unaware, as long as they would be rationally motivated by them if aware of them. This is in keeping with the fact that I can discover what reasons I have, or someone can point them out to me, although, of course, I must be aware of them in some way in order to be motivated by them. That I have a reason does not mean that I must act on it or be deemed irrational. I must have some motivation to so act, but it need not be sufficient to produce the action or relevant intention. I may have reasons for several incompatible actions none of which is rationally required of me. To be motivated is simply to have some desire or inclination toward some action, to be disposed to act in the absence of other equal or stronger reasons.

Finally, by way of clarification, the antecedent clauses that refer to being F-minded typically refer to an agent's values, her broader, more stable desires or concerns. These will typically stand in hierarchical relations to more specific desires. I am motivated to go to the Sports Authority store only if I am tennis-minded, but my concern for tennis may be based on a desire to compete or exercise, the latter desire based on a concern for my weight, and this desire based on my being health-minded and ultimately concerned for my own welfare. Values or broader concerns may give rise to more specific desires when certain states of affairs are recognized as opportunities, thereby becoming reasons for the more specific desires (for example, I see an ad for a tennis racket at the Sports Authority and come to desire to go there).

If I am F-minded and therefore motivated by F reasons, then I will be motivated by the Fness of the reasons, although I may not conceive them as such. I might, for example, be moved by the needs and interests of others and therefore be morally-minded although, like Huckleberry Finn, I don't think of myself or of these concerns and reasons as moral. The first definition indicates that there can be reasons, for example moral reasons, that do not motivate. Nor will agents necessarily be irrational for not being motivated by them. Moral reasons will motivate only as long as people are morally-minded as well as rational. The second definition indicates that subjects have moral or aesthetic reasons only if they are morally- or aesthetically-minded or concerned. The distinction between these definitions means that we can be externalists (in at least an attenuated sense) about what reasons there are and internalists about what reasons agents have. The externalism is attenuated because the analysis still makes reference to motivation in two places, and because we would not recognize any reasons as such unless at least some people were

relevantly concerned. But the distinction remains significant in that it will allow us to accommodate some of the premises in the arguments against internalism without accepting their conclusion.

Before addressing these arguments, more groundwork is needed on the nature of reasons and on what rationality itself requires of us. Just as I am understanding internalism in a strong sense here, in a sense that makes it most vulnerable to counterexamples if there are any, so I am understanding reasons in a strong sense, in the sense that connects them most closely with rationality. If acting on reasons makes one rational, this is because they both explain and justify, at least to some degree, one's actions. Reasons are what motivate and hence explain our actions, and they motivate precisely by justifying our actions in our own eyes. This is not to deny that there is a weaker sense of reason in which reasons explain without justifying. But this is the sense in which there are reasons for nonsentient events and not only actions—the reason the shuttle crashed was a hole in its wing. A reason in this weak sense is simply a cause. This is not the sense that connects essentially with rationality, our concern here. There can, of course, also be justifying reasons that are not acted on and therefore are not explanatory of actions. So justification and explanation can come apart, but our interest lies in cases of rational actions where they go together.

Reasons, then, are what motivate us, what we deliberate about, and what, often after deliberation, explain and justify our actions. As such, reasons are states of affairs or facts, such as a tennis racket's being on sale or its raining outside, that count for or against certain actions. The more standard view is that reasons are complexes of beliefs and desires.⁵ On this view, the reason I go to Sports Authority is my desire for a new tennis racket and my belief that I can pick one up on sale there. The plausibility of this account derives first from the fact that this desire and belief can indeed explain my action, being parts of the causal chain that lead to it. The fact that the racket is on sale is also part of that causal chain, having given rise to the relevant belief when apprehended via the newspaper ad, and having risen to the status of reason by relating to my desires. Which part of the causal chain is cited in an explanation is a pragmatic affair. Always only part of a causal chain will be cited in an explanation, that part on which interest is focused. Thus, externalists are right that we do not necessarily need to cite additional elements when citing factual reasons (or beliefs) as explanations, but wrong if they think this implies that desires are not also necessary parts of the relevant causal chains. And old-fashioned internalists are also right that we can explain by citing desires and beliefs, other parts of the causal chain. Beliefs and desires, instead of facts, are more likely to enter into third party explanations when the

⁵ Although the view I am adopting is also held by Joseph Raz, *op. cit.*, and by Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000).

beliefs are deemed false or the desires are for things considered undesirable. But while such facts or states of affairs as that a racket is on sale enter into deliberations as their objects and justify actions given the relevant desires and beliefs, we do not deliberate about our beliefs and desires, and these in themselves do not justify our actions. If I wake in the morning and desire to go back to sleep, I still may have no reason to do so and many reasons not to do so, given my concerns and busy schedule for the day. I could not justify additional sleep to myself or to others simply by pointing out that I felt a desire to go back to sleep.

But don't desires sometimes justify actions—isn't my desire for chocolate ice cream justification enough for my choosing it as dessert? My desire or state of being motivated is not my reason for choice, i.e. what motivates me. If my choice needs a justification, if I have any reason for my choice, I take it that the pleasant taste or its prospect is my reason. If this particular ice cream does not in fact taste pleasant to me, then I lack a reason. My bare desire, if based on a false belief, does not provide a reason. It is just as well (and in this case, given opposing considerations, better) to rid myself of such desires as to act on them.⁶ Even without false beliefs, desires in themselves do not justify or provide reasons. "Just because I wanted to" is not so much the provision of a reason as an admission that I had none. Such actions without reasons are not irrational unless incoherent with other desires and concerns; in contexts in which we have no reason for doing one thing rather than another, we need no justification for doing either. Broader concerns may also seem to provide in themselves justificatory reasons, as when I say that concern for my own welfare was the reason I backed out of the contract that had become disadvantageous to me. But here again it is more proper to say that the fact that the contract had become disadvantageous to me was my reason, given my concern for my own welfare. The concern is a background pre-condition for the reason, not the reason itself.

As for beliefs, it may seem that it is precisely in the context of false belief, belief to which no fact corresponds, that beliefs in themselves justify actions. If I falsely believe that the ice cream will have a pleasant taste, then there is no fact to justify my choosing it, but the choice itself may nevertheless appear to be justified by the belief itself when combined with the relevant desire. Again I prefer to say that I had no reason in that context for my choice, although I thought I had (and perhaps could be excused for thinking so). This preference in vocabulary is again supported by the strong sense of reason with which I am operating, the sense in which reasons objectively justify and explain. And this strong sense is justified by its giving the opponents of internalism a fair run for their money. If all reasons were desires, or desires plus beliefs, then the internalist's position would be conceptually

⁶ Dancy makes this point, *op. cit.*, p. 39.

established. That reasons are facts or states of affairs leaves it open whether they may be independent of desires or pre-existing motivational states.

If reasons are what we deliberate about and what both explain and justify our actions, then reasons are not desires or desires plus beliefs. And if desires are states of being motivated, then they are not *what* motivates (again reasons).⁷ Moral reasons, for example, are not desires and beliefs, but plausibly include consequences of actions (states of affairs that result) and types of actions (e.g. the fact that an action is the fulfillment of a promise). This leaves it open too, however, that reasons always reflect desires or motivational states, that deliberation is partly a process of self-discovery. Desires as well as beliefs are necessary pre-conditions of reasons without being reasons themselves.⁸ That it is raining outside is my reason for taking an umbrella, but only relative to my desire to stay dry, and it motivates me only given my recognition or true belief that it is raining.

A final word (for now) about the nature of desire. I have endorsed the view that desires are states of being motivated, primarily inclinations or dispositions to act or choose (perhaps only under certain descriptions or representations). This may seem to be counterexemplified by desires for things one cannot influence or choose. I want the Dolphins to win the Super Bowl next year, but I am not motivated to do anything to help them, being powerless to do so at reasonable cost. There are two possible replies. One is that, if I do have this desire, then I would be motivated to satisfy it if I were able to do so. I prefer to say that this is not a desire proper, but simply a want or wish. The distinction is supported by the oddness of “I desire that the Dolphins win,” as opposed to “I want the Dolphins to win” or “I wish the Dolphins would win”. “I desire to go to the Dolphins game,” by contrast, is perfectly natural, as that is something within my power to do. It will be necessary later, nevertheless, to distinguish more explicitly between narrower and broader senses of desire proper.

II. Rationality and Deliberation

Since some purported examples of desire independent reasons appeal to reasons that rationality in itself requires us to have, we need as a final preliminary to sketch in broader terms an internalist’s account of what rationality does require. We can then see whether the alleged counterexamples escape the terms of this description or whether they can be captured by it. If our initial definitions of reasons and of having reasons are to stand, then we cannot, on pain of obvious circularity, define rationality in terms of responding to rea-

⁷ Compare Alfred Mele, *Motivation and Agency* (Oxford: Oxford University Press, 2003), p. 16.

⁸ Philip Petit and Michael Smith, “Backgrounding Desire,” *Philosophical Review*, 99 (1990): 565-92, also relegate desires to the background of deliberation, but they continue to view desires as reasons or parts of reasons.

sons. Instead, the nature of rationality determines what reasons are, what counts as reasons. For the internalist, what rationality demands in a nutshell is coherence. Coherence, especially with the *constitutive aims* of belief and action, determines what counts as reasons. If coherence of the sort that the internalist can require can account for the rational requirement to recognize reasons proposed as counterexamples to his account, then they are not genuine counterexamples.

Let us begin here, as others have for purposes of comparison, with the clearer case of belief. The constitutive aim of belief, as others point out, is truth—beliefs are essentially states that aim at truth. To believe is to believe true. Reasons for beliefs are therefore indications of truth, of states of affairs to which the beliefs correspond. Demands of coherence account for logical reasons. If I believe that *p* implies *q*, then I cannot coherently believe that *p* and not-*q*, because then I would believe what I believe is not true. In other words, my belief would be self-defeating in terms of its constitutive aim. Incoherence is practical self-defeat, even when it comes to belief as opposed to action. Logical rationality is part of practical rationality. It is often said that we cannot have reasons to be logically rational, but in fact we can have, for example, prudential reasons to be so. What we cannot have are reasons to be rational tout court. Avoidance of incoherence or self-defeat is not a reason to be rational: it is what it is to be rational. We need not add “If a subject is truth-minded” to the definition of logical reasons, since, as noted, to believe is to be truth-minded, and we are wired to acquire beliefs. Thus, coherence or rationality in itself accounts for logical reasons. It also determines the requirement to base empirical beliefs on sensory and inductive evidence, once more in order to fulfill the practical aim of achieving truth.

The question of a constitutive aim for action that parallels truth as the constitutive aim of belief must be more difficult, since there has been no convergence on an answer. J. David Velleman, who raised the issue and addressed it most thoroughly, has taken the constitutive aim of action to be autonomy or self-knowledge, that which distinguishes full-blooded action from mere animal behavior,⁹ while John Searle denies that there is any end of action analogous to truth as the end of belief.¹⁰ Velleman’s candidates cannot be correct, since a constitutive aim for any state *S* both sets the criterion for success of *S* and determines what counts as a reason for *S*. Autonomy and self-knowledge are not criteria for successful actions. Autonomous actions are not necessarily successful, as true beliefs are, and we can gain self-knowledge from our failures as well as from our successes, fail while knowing exactly what we are doing. Nor does self-knowledge or autonomy determine what

⁹ J. David Velleman, *The Possibility of Practical Reason* (Oxford: Clarendon Press, 2000), pp. 26, 193.

¹⁰ John Searle, *Rationality in Action* (Cambridge, MA: MIT Press, 2001), p. 137.

counts as a reason. We can self-knowingly act on impulse and not reason, and while Kantians do not see such action as autonomous, the point is debatable. What distinguishes human action from that of other higher animals is not its constitutive aim, as Velleman holds, but the way that aim, if it exists, is often satisfied, i.e. with self-awareness.

Is there then, as Searle claims, nothing that stands to action as truth stands to belief? Put in positive terms, is there anything generic that constitutes what counts as success in action, toward which action as such aims? The positive answer to this question is that action aims at the fulfillment of those motivational states that prompt it or of those ends that motivate it. The constitutive aim of action is the satisfaction of desires broadly construed as states of being motivated. If we believe, we are concerned about truth; if we act, we are concerned about satisfying the desires that prompt the actions. This is so even if, as claimed earlier, we are focused on the objects of our desires and not on the desires themselves. Achieving the object is fulfilling the desire, so that if one aims at the object, one is indirectly, if not consciously, concerned to fulfill the desire. This criterion of success for action does not distinguish the merely intentional type from the fully autonomous and reflective type, but then neither does the criterion of success for belief distinguish mere belief from its fully critical and reflective version. Reflection in both realms may increase the chances of success, but it is not necessary for success. It might be objected to the analogy that motivational states specify very different sorts of ends, and that there is therefore no single internal end of action as truth is the aim of belief. But truth too requires different satisfaction conditions for different propositions believed, and different reasons indicate those conditions.

Constitutive aims, we noted, determine what counts as reasons. The constitutive aim of action is the satisfaction of the motivational states that lead to it. If this is so, then reasons indicate ways to satisfy these states: they count as reasons only relative to existing motivations. We have arrived at the internalist's thesis via a different route (which is why this is the internalist's account of rationality). But just as rationality or coherence can require certain beliefs given other beliefs or evidence, so it can require certain desires or motivational states given other desires, facts, and broader concerns. A belief in itself is not a reason to believe it true, and a motivation is not a reason to be motivated or to act, but beliefs can require other beliefs and desires or concerns can require other desires. We can perhaps without fault desire incompatible states of affairs—I might desire both to be rich and to be a philosophy professor. The former desire, however, might again be a mere wish or want (as distinguished earlier) and not a desire proper, if I have no real inclination to act on it. In any case, at the stage when desires must give rise to intentions, the demand for coherence becomes real. And that stage is itself

demanding by rationality: one must form an intention when one apprehends the conclusive weight of reasons to lie on the side of a particular action. And one must act on the intention. Otherwise, one defeats the overall aim of one's motivational states at the time. Likewise, and equally well established, one must desire and intend some means to the ends one desires and intends; otherwise, one will defeat the latter desires and intentions when one desires that they not be defeated. Likewise, if one has general concerns that one intends to satisfy, one must specify narrower desires and intentions whose satisfactions constitute ways of satisfying the broader concerns. If one intends a lucrative career, then one must settle on medicine or law or business (but not philosophy or social work).¹¹ The internalist, as opposed to the narrower Humean who denies any rational requirements on ends, need have no problem with any of this, as long as motivational states are required only by other pre-existing ones.

Irrational actions and beliefs are then those that are incoherent with one's own practical aims or ends. Of course not all false beliefs are irrational (although inconsistent with truth), but only those inconsistent with the *aim* of achieving true belief, for example those that fly in the face of known evidence or that are inconsistent with other beliefs more firmly established. Likewise, not all failed actions are irrational, but only those all of whose motives are known to be unsatisfiable, or those that defeat ends with higher expected payoffs. Incoherence in the form of self-defeat can result from incoherence among desires or motivational states themselves, from incoherence among desires or intentions and actions, or among desires and beliefs. Intransitive preferences exemplify the first of these forms, incoherence among motivational states, since satisfying any two of these preferences will defeat a third with higher expected utility than one of the first two.

Weakness of will, failure to adopt means to preferred ends, or adoption of means that defeat more highly preferred ends exemplify the second form of irrationality, incoherence between desires or intentions and actions. There is some question of how this second form of irrationality is possible, if preferences or orderings among desires are themselves established by patterns of choice or actions. But weakness of will is not only introspectively evidenced, but can be apparent from patterns of actions themselves, or from discrepancies between verbally expressed value judgments and patterns of action or choice. The would be dieter who expresses the desire to be thinner and then alternates between starvation and eating binges is an obvious example to himself and others, although if he instead consistently overeats, he may well be a victim of self-deception about his true motives rather than weak-willed. Occasional or intermittent discrepancy between verbally expressed and behav-

¹¹ For full description of this process of specification, see Henry Richardson, *Practical Reasoning about Final Ends* (Cambridge: Cambridge University Press, 1994).

iorally revealed preferences (or desire orderings) evidences weakness of will, while a more consistent discrepancy evidences self-deception.¹² Self-deception is a form of irrationality in belief, often about one's own motives, a failure to base belief on overwhelming evidence,¹³ while weakness of will is irrationality in action, but the evidence for one or the other is sometimes only subtly different.

The important point for us is once more that all these forms of irrationality exemplify incoherence in the form of practical self-defeat, indeed defeat of the constitutive aim of belief or action (if action at any given time is interpreted as aiming to fulfill one's highest preference or motivation at that time). Our question is whether purported examples of desire independent reasons we have or must rationally have can be explained instead as demanded by this requirement of coherence with pre-existing motivational states (or in some cases dismissed). For the internalist, reasons reflect our desires, and strongest reasons reflect highest preferences as prioritized in deliberation. Are there other reasons revealed in the course of rational deliberation independent of this set of coherent motivational states?

On the internalist view, practical deliberation always aims at increased coherence among motivational states, intentions, and actions, or at avoiding the types of incoherence illustrated above. Finding the best means to an accepted end, for example, is attempting to make one's actions cohere with one's intentions by finding the action that best coheres with this and other ends that one might have. When deliberating, I try to prioritize reasons reflecting my concerns and to adopt plans that will serve several at once, tying them together and removing conflicts by building exceptive clauses into those deemed to be of lower priority. This is not a matter of introspection aimed at determining which desires are stronger. Talk of strength of desires, certainly in the broad sense in which desires encompass all motivational states, is simply shorthand for priorities among reasons, ends, or actions, or for which states of affairs or objects are considered more important or valuable.¹⁴ One weighs reasons to determine priorities, not desires directly. Thus, in deciding whether to accept a new job offer, I weigh the facts that the offering department will stimulate and facilitate my research, that the campus is near a city with major cultural attractions, that the town has excellent tennis facilities but no Thai restaurants, that there is little housing within my

¹² I am assuming here that we are rationally required to pursue what we deem overall best, the defect in weakness of will being a failure to do this. For my arguments against any form of satisficing that would violate this requirement (but not against other forms) see *Practical Rules: When We Need Them and When We Don't* (Cambridge: Cambridge University Press, 2002), pp. 79-83..

¹³ "To believe what one knows ain't true" may define faith, as H.L. Menken claimed, but it also defines self-deception, suggesting a further definition.

¹⁴ Compare Elijah Millgram, *Practical Induction* (Cambridge, MA: Harvard University Press, 1997), p. 101.

budget, and that I will be leaving some excellent graduate students and friends but gaining better students overall. I attempt to prioritize these reasons (to remain private) by reflecting on them and on the consequences of satisfying or failing to satisfy them. My decision will be contingent on the possibility of other actions that tie into it, for example my finding a suitable house, where suitability again involves connections with various concerns and minimization of conflicts. New ends, such as the goal of buying a house, emerge in the course of such deliberation, and old concerns may evolve or disappear. For example, my interest in sports might lead me to try the golf courses near my new house, which might lead to a new interest that gradually supplants tennis in my scheme of things. For the internalist, such practical reasoning need not typically consist in finding means to given prioritized ends: the internalist need not be an instrumentalist. Indeed, there is no clear or fixed distinction between ends and means here. Means affect ends other than those they are aimed at and become ends themselves, and ends are means to other ends, some of which emerge as new opportunities arise in the course of pursuing other goals.

I have said that practical reasoning aims at coherence, but we can also, of course, seek variety in our pursuits—tennis as well as philosophy. But while diverse, these pursuits should not be inconsistent. I must remove conflicts in my plans regarding times for them, and if I do, their very different natures might facilitate my efforts at both, making for coherence after all. At any given time what I am doing should not defeat my doing what I have more reason to do by my own lights, but at most times there is not any single course of action whose supporting reasons predominate over all others. Only in deliberation over major decisions do we need consciously to prioritize reasons and thereby settle conflicts among them. In other more mundane contexts, barring weakness of will, I can assume that doing what I feel like doing is reflecting my coherent motivational states at the time. I need not summon up a second-order desire to be rational or to do what I have most reason to do, although I might do so when an immediate temptation threatens to overwhelm a more stable value judgment, when weakness of will threatens. If the point of practical rationality is to give greater coherence to life, to avoid practical self-defeat, then it is not surprising that reasons relate to prior concerns and that more specific desires tend to develop and replace old ones in relation to more stable and broader concerns. Internalism about reasons reflects this broader view of rationality. That the aim of practical rationality is coherence in one's life also explains such things as why taking account of "sunken costs" is not really irrational, and why later achievements that cap longer range efforts may be more significant than earlier ones of apparently same or greater magnitude.¹⁵

¹⁵ See Velleman, *op. cit.*, p. 62.

III. Hard Cases: Prudential Reasons

It is time to turn to the seeming counterexamples, reasons that in themselves seem to require new motivations instead of reflecting old ones. These apparent counterexamples include new perceptions of value that give rise to new desires or ends in the course of experience, motivated desires that seem to derive from the recognition of values or reasons instead of creating those values or reasons, moral reasons and obligations that seem to exist whether or not we want them or want to honor them, desires that we are rationally required to have, such as concern for our own futures, again seemingly for desire independent reasons, and obligations that we freely assume that then seem to bind us independently of continuing desires to fulfill them. Some of these will be handled by appeal to the distinction drawn at the beginning between there being reasons and subjects having them; some will be handled by recognition of broader concerns and their relation to narrower desires; and some by appeal to constraints of coherence or of rationality itself as sketched above, as requiring certain motivational states given others.

Beginning with the acquisition of new desires or ends in the course of experience, this might appear to derive from the novel perception of what is independently valuable, instead of reflecting pre-existing concerns. I try golf for the first time, having been talked into a round by my neighbor, expecting to be bored by the slow pace of the game, but instead find it very pleasurable and acquire the desire to play again. Elijah Millgram, who calls such learning from experience (that, for example, golf is enjoyable) practical induction, interprets the process as revealing what independently matters, and hence what one should care about. I am here learning about a valuable activity, not about myself.¹⁶ I have reason to play golf because it is enjoyable, or involves exercising one's athletic skills in a beautiful setting, and this explains my desire to play rather than the other way round.

But surely the fact that golf involves exercising athletic skills in a beautiful setting is a reason for me to play it because I find that sort of activity enjoyable (Eddy Zemach would consider it torture). And I find it enjoyable, it is the sort of thing that pleases me, because of my broader dispositions, inclinations, or motivations, for example my proclivity for outdoor sports. We noted earlier that desires and priorities or relative strengths among them are typically revealed only by what we find valuable, where valuable features are most often taken to be objective features of the objects that have them. Nevertheless, what we are really learning about when we discover value, according to the plausible premise for internalism with which this paper began, are relational properties, relations between objects or states of affairs and our motivational dispositions. Millgram in effect acknowledges this when he draws a comparison to the perception of secondary qualities, point-

¹⁶ Millgram, *Practical Induction*, pp. 80-81.

ing out that we learn something about an object when we see that it is red, not something about ourselves.¹⁷ But again, once properly conceived, we realize that in perceiving colors we learn about relational properties, however objective they may appear. Millgram is, of course, right that only perception directed outward, and not introspection, will tell us the colors of things, and he is also right that only practical induction, and not introspection, will tell us what is valuable or worth pursuing. But this need not bother the internalist, since it is compatible with our initial perceptions of value depending on our broader concerns or motivational sets, although the reasons acquired in these new experiences explain the acquisitions of new more specific desires. A motivated desire, such as my newly acquired desire to play golf, exists because of the recognition of a reason, but the reason has that status because of broader concerns or inclinations. (I am not appealing here, however, to a standing desire for pleasure; I accept that pleasure in its broad sense is simply a byproduct or indication of other value.)

I have been assuming different levels of generality in desires or motivational states, calling the broader ones concerns.¹⁸ This will parallel a similar distinction to be drawn below between broader and narrower concepts of self-interest. Desires typically nest in hierarchical structures. My concern for my welfare spawns a concern for my health, which spawns an interest in exercise, which spawns a desire to play tennis, which spawns a desire to have a good racket, which spawns a desire to go to the Sports Authority store, which spawns a desire to avoid the day's rush hour traffic, and so on. Desires in the broad sense usually do not involve negative sensations or pleasant thoughts. My concern for my welfare does not encompass yearning feelings or delightful images of satisfying it. Nor does my motivation to go to the dentist or to perform other necessary means that are of value only as means. Narrower desires for more specific objects may well involve these sensations and pleasant thoughts. Focusing only on the narrow sense underlies some arguments against internalism. Joseph Raz, for example, points out that we often do what we don't want to do (but have to), and he takes this to show that we often act for desire independent reasons.¹⁹ But the premise is true only in the narrow sense of want or desire. When we focus on such broad concerns as the concern for our own welfare, it becomes not implausible that we begin life with a set of such broad dispositions that evolve and generate more specific concerns in light of experiential input. As opposed to the golf example, it might seem that a concern for others' welfare first develops by sympathizing with the plights of others, where one is not inclined to view the value of

¹⁷ Ibid, pp. 81-82.

¹⁸ This usage follows Simon Blackburn, *Ruling Passions* (Oxford: Oxford University Press, 1998), p. 123.

¹⁹ Raz, op. cit., p. 110.

their welfare as depending on one's pro-attitude toward it.²⁰ But again, one presumably would not sympathize unless there was an initial, in this case biologically wired, disposition to do so.

Focusing on the broad sense, however, raises other objections. First, relating to our earlier discussion of irrationality, if we always do what we are motivated to do in the broad sense, if in fact our motivations are posited on the basis of our choices, then it becomes problematic how we ever act irrationally. Just as a broad sense of self-interest (to be introduced below) seems to leave no room for self-sacrifice, so a broad sense of desire seems to leave no room for irrationality in action. But we have noted that desire in the broad sense, despite lacking introspectible sensations, is not fixed by choice in action alone, but also by sincere verbal expression, indicating how the object of desire is conceptualized. When choice and sincere verbal expression come apart, we have either weakness of will or self-deception. Our actions and beliefs can fail to cohere with our evaluations of reasons, and then we have irrationality. And, indeed, these two forms of irrationality can sometimes be revealed in patterns of choice or action alone, as in the cases of the dieter-binge-eater and of the loyal spouse whose actions reveal ignorance of escapades plainly visible to all others.

A second problem with appeal to broad motivations in defense of internalism has been raised by G. F. Schueler. He claims that the internalist thesis becomes trivial once it appeals to the broad notion of desire. Since an agent will have a pro-attitude to some degree toward any action undertaken (even one displaying weakness of will or one undertaken as a means only), no reason for said action can ever be ruled out by the internalist thesis. Even if a reason seems unrelated to prior motives, if an agent acts on it, then she must have been positively disposed to do so.²¹ From the point of view of deliberation, since we often do not know what moves us until we deliberate, but know that we desire in the broad sense whatever moves us after deliberation, the internalist's thesis becomes indistinguishable from the externalist's.²²

One minor problem with this argument is that Schueler makes the common mistake of confusing desire, or the state of being motivated, with reasons, or what motivates. The internalist thesis is indeed trivial, as noted above, if all reasons consist in desires themselves. But since, even without the equation of reasons with desires, we will still always posit pro-attitudes toward the reasons that motivate us, this problem in this context does not nullify Schueler's argument. The argument does not depend on this equation. More seriously, although Schueler's point applies from the viewpoint of deliberation or of the deliberator, that is not the only, or even the principal,

²⁰ This objection was raised by Stephen Darwall.

²¹ G.F. Schueler, *Desire* (Cambridge, MA: MIT Press, 1995), pp. 55-6.

²² *Ibid.*, p. 156.

viewpoint from which this discussion is being conducted. Our question is the philosophical one of whether agents can have reasons that do not connect to their existing concerns. And that question has great practical importance not for deliberators who seek a formula to guide them, but instead for those trying to convince or provide them with reasons. It is crucial, for example, to know whether one can argue someone who is morally unconcerned into being motivated by particular moral reasons, or whether conversion in such a case requires a more basic personality change.

So far in this section we have concluded that new motives acquired via the recognition of new reasons in novel experience do not refute the internalist thesis, and that this conclusion is not trivial for appealing to broad concerns. One can discover that things have value or provide reasons for one when one had not conceived of them as such before, but these discoveries still depend on one's broader motivational sets. The next category of purported desire independent reasons appeals to reasons that rationality itself imposes upon us. We can safely skip here the requirement of willing some means to ends that we will, since the reason we must do so in particular cases clearly connects to ends and desires that we already have. The two central categories of contention here are prudential reasons and moral reasons. So first, are we rationally required to be concerned for our own futures, and if we always have reason to be so concerned, is this an example of a desire independent reason?

The question about concern for one's future is not really a single question admitting of a single answer, since it masks several relevant distinctions. First, there is the distinction between narrow and broad self-interest, where the satisfaction of the former but not the latter makes essential reference to oneself. If I take an interest in others' welfare, then their welfare is part of my broad self-interest, although I can sacrifice my narrow self-interest to promote their welfare, for example by spending money on them instead of on myself. What satisfies any of my concerns contributes to my broad self-interest, whereas my narrow self-interest reflects only my concern for myself. Second, there is the distinction between the future satisfaction of desires I have now and the satisfaction of desires that I know I will have but do not have now. And third, there is the distinction between satisfactions I will enjoy at a later time and those I might enjoy at an earlier time or now.

When it is said, for example by Stephen Darwall,²³ that I need have no concern at all for my own welfare or own future, this is true in the narrow sense of self-interest. My narrow self-interest, or the fact that something would serve it, may give me no reason to pursue that thing. Or more precisely stated, I may without rational fault have no desires for things or states of affairs that make essential reference to me, for example that I enjoy some-

²³ Stephen Darwall, *Welfare and Rational Care* (Princeton: Princeton University Press, 2002), p. 3.

thing or own something. If I lack any such desires, I simply have no narrow self-interest. Certainly such lack would be highly unusual and probably contrary to biological instinct, but, for all that, I might care only about art, or the revolution, or God. And if that is the case, then there appear to be no desire independent reasons for narrow self-concern, the ignoring of which is a rational fault. Of course, I might have moral or revolutionary or religious reasons to be concerned for my own survival, but then, since I would be morally or revolutionarily or religiously-minded, my motivated prudential reasons would pose no threat to the internalist. I can be required only by my other concerns to have such narrow prudential concerns.

Is one rationally required to be concerned with one's broad self-interest, and is the reason to be so concerned desire independent? Yes and No. If one has any desires at all, then one is concerned for one's broad self-interest, for that is what it is to be so concerned.²⁴ I cannot be unmotivated to act on my own desires, for desiring just is being in a motivational state. If I seem, for example, too depressed to be motivated to satisfy my desires, my depression is really causing me to lose or temporarily lack these desires. Certainly it is difficult to imagine a person with no desires, a person who did not perceive any states of affairs or prospective states of affairs as better or worse, or as reasons, and such a person in any case would not be a rational agent. So every rational agent who has desires and reasons also thereby has a concern, and a reason for concern, for her broad self-interest. This reason is an internal or desire dependent reason. A similar argument applies to concern for the future satisfaction of present desires (if one does not know that one will come to lack these desires before they can be satisfied). Of course, all present desires, if not already satisfied (and therefore no longer desires?), must be satisfied in the future if at all. So having any desires or concerns is having a concern for the future in which they could be fulfilled. So having any reasons that reflect those concerns is having a reason to be concerned about the future, an internal or desire dependent reason.

A more difficult question relates to reason for concern for satisfying predicted future desires that I do not have now. Am I rationally required, for example, to avoid, or at least to be negatively motivated toward, something that will damage my health in twenty years, when I am presently unconcerned about my health then? In order to isolate a desire independent reason, we must first eliminate any present desires that may await satisfaction at that later time and for which my survival or health might be necessary, for example a concern that my children still have a father in twenty years. Having done that,

²⁴ At least in a transparent sense of concern. Again, I believe there are both transparent and opaque senses. In the former sense, if I am tennis-minded, then I am sports-minded, even if I do not know that tennis is a sport. In the latter sense, I must have a concept of what I desire, and I desire things only under certain descriptions.

there is still the question whether I can consistently take an interest in my health now but not in my future health, while foreseeing no relevantly changed circumstances or change in attitude, that is, knowing that I will continue to care about my health later. Since it will still be the same self with the same concern, this would be like being concerned for my left side but not for my right side without seeing any reason for distinguishing these spatial parts of myself. It seems to me that if there is a rational requirement or reason to be concerned for my future self when I lack that concern now, it is because of some such demand for coherence with my present concerns for myself. If I am concerned about present pain, for example, then I must be concerned about future pain if I foresee no change in attitude, since that future pain will be present pain to me when its time comes. On the other hand, if I simply predict a future concern based on information about the typical concerns of older people, then there is no rational demand to prepare for its future satisfaction now. If that is correct, then the requirement to be concerned for my future self once more need not concern the internalist. We have failed again to produce a desire independent reason.

A final issue in regard to prudence is the irrationality of choosing a lesser satisfaction now over a greater one later (discounting for the probability of the latter's not occurring). This once more is simply weakness of will or failure to act on one's strongest reasons once they have been prioritized. The irrationality of weakness of will derives from the rational requirement to do what is overall best (in terms of broad self-interest) at a given time (although we noted that often, especially in less important matters, no reasons clearly override or determine an overall best). To prioritize our reasons or values and then sacrifice the higher or stronger to the lower or weaker is to be incoherent. It is to defeat one's own preference. But once more there is no sharp line to be drawn between the requirement as it applies at a given time and over time. The same self is extended in time, as are temptations and greater satisfactions. The success of one's ongoing projects is not momentary, and in general, the more extended the project, the greater the satisfaction at its success. Our point is once more that that the requirement is one of coherence among existing motivational states and actions or choices.

IV. Hard Cases: Moral Reasons

We turn finally to the main concern and bone of contention for the critics of internalism as we are understanding it: moral reasons. Those who hold that moral judgments or reasons must be intrinsically motivating are called internalists in ethics, but they are judgment internalists, not reason internalists as we are understanding that position. They hold that the recognition of moral reasons (or moral judgments) must motivate rational agents whatever their existing contingent desires. According to them, the moral reasons that agents

have are not limited by their other desires or motivational states. As we have defined it, these are reason externalists, and the internalist must refute their arguments. Obviously I cannot in the remaining space here refute while doing justice to every neo-Kantian and other argument for the intrinsically motivating force of moral reasons or judgments. I shall instead briefly consider some of the most prominent recent arguments and show why acceptance of points previously made here reveal questionable premises or inferences in them.

Let us begin with Michael Smith. His main argument to the conclusion that we are rationally required to be motivated by moral reasons is not the one most widely criticized, as he himself makes clear,²⁵ but a much simpler and more plausible one with much in common with what has been said here previously. It too employs a notion of rationality as coherence and a notion of a normative reason as motivating rational agents. It follows directly from Smith's concepts of reason and rationality that an agent who believes there is a reason, specifically a reason with moral content, must, if he is rational, be motivated to some degree to act on that reason. He holds first that to believe that some action is right is to believe that there is a normative moral reason to do it. A reason for him is anything that would motivate a fully rational agent. If the content of the belief that there is a reason is that there is a consideration that would motivate one's fully rational self, then one must, if one is rational, be motivated by that consideration.²⁶ The inference is as simple and straightforward as can be.

In replying, we may ignore the question whether the analysis of a reason here, which is close to mine, captures the content of everyman's concept of a reason and therefore may be substituted into belief contexts. Setting that question aside, the argument, as thus simply stated, ignores two points made earlier here: first, that there is a crucial distinction between there being a reason and an agent's having a reason; and second, that a seeming reason based on a false belief is not a reason. To believe that an action is right or morally required is indeed to believe that there is a moral reason to do it, but, given our initial definitions, that only gives an agent a reason if she is morally-minded. Similarly, to believe that the color of one's shirt clashes with that of one's pants is to believe that there is an aesthetic reason against wearing them together, but that gives one a reason only if one is aesthetically-minded. To believe that someone has done one a favor is to believe that one is required by etiquette to thank her, but that gives one a reason only if one is

²⁵ Michael Smith, "In Defense of 'The Moral Problem'," *Ethics*, 108 (1997): 84-119, p. 116.

²⁶ Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), pp. 180-87. This main argument has been strangely neglected in the literature critical of Smith's book. A secondary argument, presented earlier in the book and to be addressed here shortly, has received far more critical attention.

concerned about etiquette (of course there may be reasons of other sorts to be so concerned).

In an article criticizing Smith, David Brink comes close to drawing the distinction I have drawn, except that he distinguishes instead between two types of reason. He appeals not to the common distinction between normative and motivating or explanatory reasons, but to a distinction between those reasons that simply express social or behavioral norms and those that it would be genuinely irrational to ignore.²⁷ According to him, Smith fails to show that moral reasons fall into the second class. Smith replies by distinguishing requirements of morality from those of etiquette. The former, according to him, have a content that is not arbitrary, while the contents of the latter are arbitrary or merely conventional and therefore do not amount to rational requirements.²⁸ But aesthetic reasons and reasons of etiquette are not arbitrary. It is not arbitrary that certain colors clash and others do not, and it is not arbitrary that when someone does you a favor, you thank him instead of calling him a horse's ass. This is not to deny some elements of pure convention in both domains, as there well may be in the domain of morality as well. But in our example, it is not conventional that calling someone a horse's ass is insulting him and that one does not insult a benefactor in response to his help. Yet we are not required on pain of irrationality to be motivated by these reasons of etiquette. There need be no incoherence involved in lacking these concerns (although there may be, given one's other concerns). Similarly, if beliefs about moral rightness are beliefs about what moral reasons there are and not necessarily about what reasons one has, then there need be no incoherence about judging certain actions right but being unmotivated to perform them.

One might object to my use of the distinction here that one cannot escape from one's moral obligations by being morally unconcerned. That is true, but the same is true of one's aesthetic and politeness obligations. The response, "I don't care about that," does not falsify the criticism that I should have thanked my benefactor or should not have worn the shirt with the pants I am wearing. But whether one's moral or aesthetic or politeness obligations give one reasons does depend on one's concerns. Thus, I am maintaining, the existence of a certain type of moral reason might constitute a moral obligation, but whether I have a reason to fulfill a moral obligation that applies to my behavior depends on whether I have moral or other concerns that require me to have moral concerns. Legal obligations are similar. Whether I have a rational obligation or reason to follow Jewish law and the obligations it imposes on

²⁷ David Brink, "Moral Motivation," *Ethics*, 108 (1997): 4-32, p. 20. My distinction makes it clearer that whether it would be irrational to ignore certain reasons depends on one's concerns. And it gives a neater analysis of reasons.

²⁸ Smith, "In Defense of 'The Moral Problem'," pp. 109-10.

me as a Jew depends on whether I have religious concerns (or perhaps family concerns that require me to care about my religion).²⁹

There is also the problem for Smith's argument of false belief. As noted earlier, in my view if I believe I have a reason R to A, but R seems to be a reason only because of a false belief, then I falsely believe I have a reason R. If I have no reason, am I rationally required to be motivated by my mistaken belief that I have a reason? Reflecting on the character of Huckleberry Finn provides an answer in the moral realm inimical to Smith's position. Huck falsely believes that morality requires him to return Jim to his slave owner. Being motivated by his friendship for Jim and by sympathy for Jim's plight, he vows to renounce his moral concerns and face damnation. Certainly Mark Twain did not intend us to see Huck as irrational in failing to be motivated by what he perceived to be a moral requirement or strong moral reason, and I doubt that any readers have ever interpreted the character that way. It is not that Huck's concern for Jim overrides his motivation to follow his moral judgment: having perceived in his own mind what morality requires, Huck will have none of it and is not in the least irrational in this attitude toward the perceived requirement.

Smith himself should at least be tempted by this normal interpretation of the novel in so far as he holds that overriding a direct concern for right making properties by a concern to be moral is fetishistic, involves, in Williams' famous phrase, "one thought too many". He, however, utilizes this point in a second more widely criticized argument for the claim that moral judgments are intrinsically motivating.³⁰ According to him, only such intrinsic motivational force can explain why a virtuous person's motivations reliably track changes in his moral judgments. Absent such force, a person might continue to care more about properties she previously thought to be right-making even after changing her moral judgment regarding them. Or, if she reliably does change her concerns to match her moral judgments, it could only be because of the fetishistic desire to be moral.³¹ But the Huck Finn case shows that a virtuous and rational person's motivations do not necessarily reliably track his moral judgments, as they would if such judgments were intrinsically motivating on pain of irrationality. And to the extent that motivations do track moral judgments, Smith does not explain why, in the absence of intrinsic motivational force in the judgments, this cannot nevertheless result from direct empathic and sympathetic concerns for the needs, interests, rights, and plights of others, as much as from the desire to be moral. (When one changes

²⁹ I believe legal obligation as a citizen is similar but will not argue that thorny issue here.

³⁰ For some additional criticisms, See Russ Schafer-Landau; "Moral Judgement and Normative Reasons," *Analysis*, 59 (1999): 33-40; Hallvard Lillehammer, "Smith on Moral Fetishism," *Analysis*, 57 (1997): 187-95.

³¹ Smith, *The Moral Problem*, pp. 71-6.

one's mind about what affects these, one changes one's moral judgments and motivations *if* one is morally-minded.)

The motivations of real people do not track their moral judgments to nearly the extent they would if such judgments were intrinsically motivating. Sometimes, as in Huck Finn's case, this is good, especially where the judgments are overly legalistic or code bound; more often, it is not good. Smith's position has been confronted by the amoralist,³² who, he must claim, does not use or understand moral language as do the rest of us.³³ But real people fall somewhere between this villain and the virtuous, perfectly reliable tracker. They are motivated by some or most moral matters but not by others. I am only somewhat ashamed to admit that I am an amoralist when it comes to eating meat. I am convinced by the vegetarian's arguments but am unmotivated by them, being satisfied to treat others of my species with kindness and respect. I do not see this as weakness of will: my moral motivation in this case is not simply overwhelmed every time I am confronted by a juicy steak (being confronted by a steak feels different from being confronted by a big slice of cheesecake that I know I should resist by my own lights). And it certainly is not the case that my use of moral terms like "right" is radically ambiguous, being strangely aberrant and uncomprehending when I use them in gustatory contexts. I understand perfectly well what it is to judge that eating meat is wrong. No ordinary form of misunderstanding or irrationality, e.g. weakness of will or self-deception, seems to explain my being unmotivated by the judgment. But aren't I irrational in the sense earlier defined, i.e. my actions fail to cohere with my evaluation of the reasons? No, because I do not judge that I have overriding reasons not to eat meat, only that there are such reasons for those morally motivated by them. In my case, unlike Huck Finn's, it would be morally better if I could summon, as people sometimes can, a desire to be moral to remedy my lack of direct concern for the bovine victims. Such use of this desire is no more fetishistic than is the similar occasional use of the desire to be rational as an aid in avoiding or overcoming weakness of will.

Thus, Smith's arguments fail, and his position appears to be widely counterexemplified, at least if many people have moral failings similar to my relation to the cows. We are not motivated by certain recognized moral requirements. The other best known recent attempt to establish the claim that we are rationally required to be morally concerned is that of Christine Korsgaard. Her argument is of special interest here because she explicitly shares the main premise regarding the origin of value with which this paper began. Values originate with us sentient beings, or with us in cooperation with the world. (For her, they originate in free choice, which is more doubtful.) And

³² See, for example, Brink, *op. cit.*, pp. 21-5.

³³ Smith, *The Moral Problem*, pp. 69-70.

she again offers a strictly coherentist argument for the claim that we are rationally required to be morally motivated. If we value any reasons at all, she argues, then we must value humanity, which concern gives us moral reasons. Once more I will simply note where points defended earlier here contradict her premises. These premises can be summarized as follows:

- (1) Desires become reasons, acquire value, through reflective endorsement from the perspective of one's practical identity.³⁴ (One's practical identity—philosopher, father, tennis player, and so on—is a description under which one values oneself.³⁵ One evaluates desires by deciding whether acting on them is consistent with one's practical identity.)
- (2) Therefore we must have practical identities to act on reasons, and we must act on reasons because we are humans, reflective beings.³⁶
- (3) The value of reasons, our reflective endorsements of them, derives from the value of our practical identities, our endorsements of *them*, which derives from the value of being human.³⁷
- (4) Therefore, if we act on any reasons, we must value our own humanity.
- (5) If we value our own humanity, we must value that of others, or humanity in general, because reasons are public; their normative force is shared.³⁸

Premise 1 is completely unmotivated by, and indeed at odds with, the analysis of reasons offered here. Desires, I have emphasized, are not reasons whether endorsed or not, and what are reasons do not await endorsements to become so.³⁹ As noted, I can have reasons I am not aware of and so certainly have not endorsed as such. Korsgaard herself is not consistent in her usage here. When she speaks of particular reasons, they tend to be facts or states of affairs, such as that Susan is in trouble.⁴⁰ And the main reason for our moral concern is not itself a desire, but the fact that humanity is the source of all value. That she analyzes reasons as endorsed desires is not trivial to her full

³⁴ Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), pp. 93-7.

³⁵ *Ibid.*, p. 101.

³⁶ *Ibid.*, pp. 101, 120-23.

³⁷ *Ibid.*, p. 121.

³⁸ *Ibid.*, pp. 135-40.

³⁹ Compare Rachel Cohon, "The Roots of Reasons," *Philosophical Review*, 109 (2000): 63-85, p. 69.

⁴⁰ Korsgaard, *op. cit.*, p. 243.

argument, however: at one point she says that it is because we tend to regard our desires as reasons that we value our humanity.⁴¹ Nevertheless, this shifting usage may not be so damaging to the argument as a whole, since the fact that our desires are in some sense sources of reasons may be sufficient for the contribution that her equation of desires with reasons makes to the argument.

A more serious problem with this and the next premise is that most of our reasons appear to have nothing at all to do with our practical identities. If I have never played tennis before but think I might like to try it (I am for the moment tennis-minded), I may have a reason to go to Sports Authority to buy a racket, just as I have a reason to go to Ben & Jerry's if I'd like some chocolate ice cream.⁴² (Don't say my practical identities here are "person who sometimes likes to try new things" or "occasional ice cream eater".) The claim that we reason by checking our desires against our practical identities also presents a strangely self-centered view of deliberation, focused on ourselves instead of on the facts and prospective states of affairs that are our reasons, as described above. (My practical identity may not be self-centered, but my focusing on it is.) Korsgaard says that if I am not governed by a concept of my identity, then I have no reason to act,⁴³ and that it is because I must have reasons that I must value my humanity as the source of my identity. But if reasons for the most part do not depend on my identity, but only on my more or less specific desires, the argument in this form collapses.

An equally serious problem arises in the transition from premise 3 to 4. Korsgaard pictures a reflective regress, in which we question our (endorsements of) our particular reasons, and then question our contingent identities on which those reasons are based, until we arrive at our humanity or reflective natures, the ultimate source of all value, which is not contingent for us. But arriving at the source of value, if Korsgaard has done so, does not show why we must value that source. The assumption that we must value the source of value if the value that is effect is to be really valuable is reminiscent of the Rawlsian assumption that any causal antecedent of what counts as our criteria of desert must itself be deserved, if the criteria are to be endorsed as basic. But our basic moral convictions do not in fact dismiss productive effort, for example, as deserving of reward simply because it may be causally dependent on factors beyond our control. Conversely, if value emerges from a particular source, it is not implied that that source must be valuable or valued by us.⁴⁴ Our contingent practical identities have as their sources not simply

⁴¹ Ibid, p. 249.

⁴² Compare G.A. Cohen, "Reason, Humanity, and the Moral Law," in *The Sources of Normativity*, p. 185.

⁴³ Korsgaard, op. cit., p. 123.

⁴⁴ Samuel Kerstein, "Korsgaard's Kantian Arguments for the Value of Humanity," *Canadian Journal of Philosophy*, 31 (2001): 23-52, p. 33, makes a similar but broader point: that if one thing confers a property on another thing, the first need not have that property.

our humanity or reflective need for reasons, but complex cultural and personal histories.⁴⁵ In my and probably your case, this history may well include a slave class in ancient Greece that made the leisure for philosophical reflection possible for others. Must we, therefore, in order to endorse our identities as philosophers, value the institution of slavery of the sort that existed then? Need we, for that matter, always value our humanity as the source of all value and reasons, if we are to have any reasons? Is suicide, therefore, never rational—can a reflective agent who is terminally ill and in pain never have unopposed reasons for contemplating it?

Finally, premise 5 is the most bewildering of any. It is puzzling first why, if it can be defended, any of the others are required or even relevant. If all reasons are really shared, this should be sufficient in itself to establish morality. Korsgaard defends it by appealing to Wittgenstein's private language argument (??) and to examples in which reasons are shared, such as two people's reasons for meeting at a certain time. But the private language argument can establish only that reasons, like sensations, are communicable, not that they are shared in the stronger sense of being held in common.⁴⁶ And the fact that people can share reasons in this stronger sense does not show that they always do so. In fact they don't. I may have a reason to go to the Sports Authority store this morning, but of course that does not give you any reason to go there or even to help me get there if I need help to do so. You can go back to sleep instead without being irrational for not providing me with transportation. Shared reasons generally depend on contingent special relations among those who thereby share them. These relations, of course, may be quite broad and common: morally-minded people share reasons not to harm one another, and so on.

In summary, what Korsgaard sees as requirements of coherence far exceeds my earlier descriptions of such requirements. You might wonder why I could not have much more briefly granted her requirements for the sake of argument and then held that, being requirements of coherence, they once more create only reasons acceptable to the strong internalist. That was the strategy employed in response to the arguments that saw prudential reasons as rationally required. But it does not work against Korsgaard, since she requires coherence not with contingent pre-existing motivational states, but with our noncontingent identities as humans or reflective beings. One might say that

I prefer to attack directly the narrower claim about the source of value being valuable, since Korsgaard needs only this narrower claim for her argument, and it might be true even if the broader one is not.

⁴⁵ This point is emphasized, although to a different purpose, by Christopher Gowans, "Practical Identities and Autonomy," *Philosophy & Phenomenological Research*, 44 (2002): 546-70.

⁴⁶ This seems an obvious point and is noted by other critics. See Allan Gibbard, "Morality as Consistency in Living," *Ethics*, 110 (1999):140-64, pp. 161-3; also Raymond Geuss, "Morality and Identity," in *The Sources of Normativity*, pp. 197-8.

in her view we must value humanity only if we are to act on other reasons and hence have other contingent desires or concerns that these reasons reflect. Then the reason to value our humanity is one acceptable to the strong internalist. But acting on reasons is itself not contingent for her or for us: it is because we have the reflective natures that we do, because we are essentially reflective beings, that we must value our humanity according to Korsgaard. Hence her argument must be refuted and not accepted by the strong internalist.

The other, earlier strategy will work against the last recent argument for desire independent reasons that I shall consider, that of John Searle. Searle argues that when we freely assume obligations by committing ourselves to future conduct, we then acquire reasons to fulfill those obligations. These, according to him, are desire independent reasons, since we have reasons to act as we are committed to acting whether or not we then want to do so. Such commitments and obligations include not only explicit contracts and promises, but also many more common types of behavior. If I order a beer in a bar, I am committed and so have reason to pay for it whether or not I desire to do so.⁴⁷ The swindler no less than the honest dealer has a reason to pay for that beer once he orders and drinks it.

But does he? Once again we must carefully distinguish cases here, and I believe we must also distinguish obligations from reasons to fulfill them, as we did above. For normally morally-minded people there is no question that freely entering into such transactions gives them reasons to hold up their end of the agreement. As noted earlier, their moral concerns include not only or primarily the *de dicto* desire to do what is right, but also concerns for honesty, fair dealing, the rights and welfare of others, and so on. And, of course, being concerned also for their own reputations and future opportunities, they also have prudential reasons to fulfill such obligations, reasons in themselves almost unopposed when it comes to such trivial burdens as paying for a beer. Morally concerned agents can commit themselves to future interactions with others in the same way that prudentially concerned agents can commit themselves to rules or patterns of conduct that resist predicted temptations. Not carrying through on such commitments without change in underlying attitude is in both cases irrational weakness of will, not acting on an intention formed and maintained with the weight of reasons behind it. These are again internal reasons, since it is a matter of coherence among motivational states and actions over time. Freely and rationally forming an intention or accepting an obligation without then acting on it is a form of practical self-defeat.

At the other end of the spectrum is once more the amoralist. Does he have reason to pay for that beer? Certainly, if he is at all normal in other ways, he has ample prudential reason. Does he have moral reasons? A beer having been

⁴⁷ John Searle, *Rationality in Action*, pp. 137-8.

ordered, there are such reasons amounting to a moral obligation to pay, but do they give him such reasons? In the previous case of the morally-minded agent, we saw that she acquires moral reasons through the voluntary acceptance of the obligation and hence of the moral norm that underlies it. Does the amoralist similarly accept an obligation to pay? For Searle he does so simply by uttering the words that order the beer. Searle points out that insincerity is possible only because most who utter these words do thereby accept an obligation, and that deception is possible only because others understand the deceiver as accepting an obligation.⁴⁸ But that successful deception depends upon such common practices and understandings among the morally-minded does not show that the amoralist really accepts an obligation, endorses a moral norm, or thereby incurs a reason. The case is similar to that of the successful liar. The liar is not committed to the truth of what she utters. She will not draw inferences on the basis of it and is not irrational not to do so, although others must understand each other as well as her as making such a commitment by uttering her words. Analogously, the amoralist only appears to accept voluntarily a commitment and thereby give himself a moral reason to comply. There are moral obligations not only to keep commitments made, but to accept commitments one appears to make. But these obligations no more give the amoralist reasons than do any other moral requirements.

The final and most difficult case again involves changes in motivations or attitudes. A change for the better is not hard to handle. If the amoralist repents, she thereby gives herself reason and commits herself to fulfill her obligation. But this is no indication that mere recognition of a moral obligation suffices to require her to be motivated by it, any more than it would be in the case of a legal or aesthetic obligation. A more fundamental change in attitude or motivational set would be necessary. The really difficult case is the highly unusual one in which a normally morally-minded person incurs a moral commitment and then loses all moral concerns. Does that person have a reason to fulfill the obligation she freely accepted? The case is not only highly unusual, to say the least, but extremely difficult to identify in practice. A more common parallel prudential case is that in which a person commits to a prudential rule, say not to snack between meals. There appears to be a distinction from the first person point of view between irrational weakness of will and rationally innocent change of mind, perhaps a reordering of priorities between gustatory pleasure and weight loss. But even from this point of view, the possibility of self-deception blurs the line. Too many changes in commitments betokens fickleness or lack of integrity if not irrationality in both moral and prudential matters. But this is once more a matter of coherence over time among concerns, intentions, and actions—an internal matter.

⁴⁸ Ibid, p. 197.

In any case, it would be crazy for the opponent of strong internalism to base his whole argument for desire independent reasons on this bizarre and problematic case.

V. Conclusion

To the morally-minded person, moral requirements seem to have intrinsic motivating force. Moral judgments are meant to be action guiding. If I express such a judgment to another, I normally intend to motivate that person; and if I deliberate myself about what is morally required, it is normally because I know I will be motivated by the outcome of my deliberation. "I ought to do A, but I have no reason to" is certainly odd, to say the least. And once I convince someone else of a moral requirement that applies to him, I should not also have to convince him that he has reason to do as he is morally required to.

These mundane points perhaps motivate the more complex arguments of philosophers for the intrinsically motivating force of moral reasons on rational agents. And the belief that moral reasons must be rational requirements mainly motivates internalism's opposition, the belief that there must be desire independent reasons. Most of the points in the previous paragraph are indeed true, but they are not decisive as premises in an argument against strong internalism. Just as moral requirements seem in themselves to supply agents with reasons, so values and secondary qualities seem to be purely objective properties of things. But philosophical reflection convinces us otherwise. When I express moral judgments to others or deliberate morally myself, I assume normal moral concerns. That explains the oddness of "I ought to do A but have no reason to" as the outcome of deliberation, and it explains why I normally don't have to convince someone that she has reason to do what she morally ought to do. But we do need to train, educate, or motivate people to be moral. Indeed, the deepest moral problem we have in society is the only partially successful outcome of such social indoctrination.⁴⁹ If this education were simply a matter of replacing ignorance with knowledge, if it were simply a matter of getting people to grasp moral truths, the problem would be far more tractable, similar to that of illiteracy. But the defect or vice of the immoral or amoral lies in their motivations or lack of concerns, not in their reason. Sociopaths need not be fools or mentally challenged, or even philosophically unsophisticated.

I have assumed as a basic premise here that values are not subject-independent. This requires the opponent of strong reason internalism to produce convincing examples of desire independent reasons. I have tried to show that

⁴⁹ This indoctrination, although not consisting in pointing out reasons that people already have, is not unjustified manipulation, since we are all better off when all accept moral reasons than when none do.

by distinguishing between there being reasons and subjects having reasons, and by determining what rationality or coherence requires and what it does not require, the internalist can accommodate or dismiss and explain away the plausibility of such supposed examples. Since what is valuable depends on our values, and since our values derive from our coherent sets of concerns or motivations, reasons, which reflect values, are also relative to those motivational states. Moral reasons are no different in this respect. The longing for an objective basis for moral requirements remains just that in the absence of objective values. Rationality fails to bridge the gap between a constructivist theory of value and a theory of obligation that posits desire independent reasons.^{50 51}

⁵⁰ I have not considered Hobbesian arguments from prudential rationality here, as I have done so in *Practical Rules: When We Need Them and When We Don't*.

⁵¹ I thank Stephen Darwall, an anonymous referee for this journal, and the philosophy discussion group at the College of William & Mary for helpful comments on earlier drafts of this paper.