Noûs

NOÛS 43:3 (2009) 395–427

Receptivity and the Will

Edward S. Hinchman
University of Wisconsin-Milwaukee

The task of a broadly internalist approach to agency is to explain how an agent's all-things-considered practical judgment has necessary implications for action. The approach faces its chief obstacle in two species of possible akratic break: between judgment and choice or intention, and between choice or intention and action. The standard internalist strategy for overcoming the obstacle is to argue that an agent's all-things-considered judgment needn't determine action in order to guide it, and that there would still be an 'internal' link between judgment and action if *global* akrasia were impossible. That would at least entail an important degree of necessitation: even if there is an akratic break in a given case, the agent couldn't *always* be akratic.[1]

It is easy to feel the force of the intuition that global akrasia is impossible—but why? If global akrasia is impossible, it seems that must be because of some feature of the individual case. In what follows I focus on the individual case directly. I argue that the two species of akratic break are not importantly different: in each case the akrasia manifests a single species of irrational self-mistrust. I aim to vindicate internalism by showing how rational agency rests on our capacity for a kind of trusting receptivity to the verdict of judgment. To call the relation *receptivity* is to characterize it as fundamentally passive. To call it *trusting* receptivity is to ensure that the passivity is not incompatible with agency, since trust retains a crucial degree of control. I'll argue that the best way to meet the externalist argument from akrasia is to abandon the assumption that the will must be a locus of activity.

I'll develop and defend the core of my account in sections I through IV and consider objections in section V. Throughout this discussion I'll assume that the relations between judgment and choice (or intention) and between choice (or intention) and action unfold sequentially—as if the relation between

© 2009 Wiley Periodicals, Inc.

judgment and action had to take the form of a process. In section VII, I'll drop that assumption and offer analyses of practical judgment, choice and intention that codify aspects of a single synchronic act. Those analyses will enable me to isolate and explain, in sections VIII and IX, the aretaic basis of judgment's rational authority, an explanation toward which I'll have cleared a path in section VI. I'll conclude, in section X, by drawing a link between that virtue-theoretic approach and an approach theorized in terms of care. Most fundamentally, I'll argue, akrasia is a failure of self-care.

I

Corresponding to the two species of akratic break we can define distinct species of internalism. We might call 'judgment internalism' the thesis that your all-things-considered judgment about what you should do bears an internal relation to your choice or intention to act accordingly.[2] And we might call 'volitional internalism' the thesis that your choice or intention bears an internal relation to such a judgment. To call these relations 'internal' is to say that they're in some sense necessary or non-empirical. We can assume that these internalist theses get determinate enough content from the way they are threatened by the possibility of what I'll call 'incontinence,' in which you judge, all things considered, that you ought to φ and yet without changing your mind proceed to choose or intend or otherwise will to do something incompatible with φing. The possibility of this first akratic break seems to entail that there is no necessary link in either direction between judging and willing: you can will contrary to your best judgment, and your best judgment can fail to engage your will. The second akratic break, which I'll call 'weakness,' occurs between an intention and its execution: you resolve to φ at some time t but then, without forgetting or changing your mind, fail to φ at t (or even attempt to). This threatens what we might call 'resolve internalism,' the thesis that intending, resolving or otherwise willing to φ bears an internal relation to actually φing (or at least attempting to).[3]

The threat in each case is that we cannot understand how such counter-normative agency is possible without positing a mediating faculty of the will with one or both of these features: to accommodate incontinence, it must function independently from judgment; to accommodate weakness, it must be exercised with greater or lesser degrees of 'strength.' Since such a faculty of the will would falsify the relevant internalist thesis, any view that posits such a faculty will count as a species of externalism.

Jay Wallace presents this rationale for externalism about the relation between judgment and intention or choice:

> Human agents have the capacity for a sophisticated kind of rational agency, insofar as they can reach independent normative conclusions about what they have reason to do, and then choose in accordance with such normative conclusions.

> This capacity presupposes that we are equipped with the power to choose independently of the desires to which we are subject. Once we have this power, however, it can be put to use in ways that are at odds with our own practical judgments about what we have reason to do. That is, we can treat our disposition to do what we ought as a further desire from which we set ourselves apart, choosing to act in a way that is at variance with our reflective better judgment. This may be regarded as a hazardous by-product of the capacity for self-determination that makes rational agency possible in the first place.[4]

Again, I'm calling this species of akratic break 'incontinence.' (Wallace calls it 'akrasia,' which I'm using as a more general term to cover both species.) Wallace's externalism opposes what I called judgment internalism and volitional internalism.

Richard Holton offers a similar pitch for externalism about the relation between intending, choosing, or more generally resolving and actually following though on the resolution:

> Sticking by one's resolutions is hard work . . . . It certainly doesn't feel as though in employing will-power one is simply letting whichever is the stronger of one's desires or intentions have its way. It rather feels as though one is actively doing something, something that requires effort. My suggestion is that effort is needed because one is actively employing one's faculty of will-power . . . [,] a kind of mental effort . . . . On this picture, then, the effort involved in employing will-power is the effort involved in refusing to reconsider one's resolutions; and the faculty of will-power is the faculty that enables one to achieve this.[5]

Again, I join Holton in calling this second akratic break—of irrationally failing to stick by a resolution—'weakness of will.' Holton's externalism opposes what I called resolve internalism.

Now here's a thumbnail sketch of the position that I believe vindicates all three species of internalism against these externalist arguments. I agree with Wallace that incontinence is a hazardous by-product of the capacity that makes rational agency possible in the first place. And I agree with Holton that avoiding weakness is by no means easy. What's not easy, on my account, is trusting reasonably in the face of the twin phenomena of unruly desires and less than perfectly trustworthy judgment or intention. The difficulty lies in reopening deliberation *when but only when* the untrustworthiness of the judging or intending self requires it. Since the question is whether to redeliberate, it must be settled through the exercise of a non-deliberative form of intelligence. I hold that this intelligence is most fundamentally a counterfactual sensitivity to evidence of untrustworthiness in the judging or intending self: if you had evidence that your judgment or intention had been untrustworthily formed you would not simply have followed through on it. I put it counterfactually, but it of course should also cover the actual case, if there is in fact evidence that you are untrustworthy. By 'evidence' I mean good but not necessarily conclusive evidence. Even inconclusive evidence of untrustworthiness can undermine trust. Insofar as you actively weigh positive evidence of trustworthiness, you act not from trust but from your own deliberative judgment—and the question remains whether to trust that judgment.[6]

This non-deliberative form of dispositional thinking is not at all easy to exercise. One often trusts oneself when one shouldn't and fails to trust oneself when one should. But reasonable trust along these lines is what makes rational agency possible: without it, we would be either compulsive redeliberators or incapable of changing our minds. What makes agency possible is thus what makes these two species of akrasia possible: the rational requirement that one stop and redeliberate when but only when one's judging or intending self is untrustworthy.

The challenge, then, lies on both sides of trust: too little mistrust, but also too much. The hazards of too little mistrust are easy to appreciate, if hard to avoid: you follow through on a foolish decision, overlook how available options have unforeseeably changed, or otherwise fail to resist flawed judgment or choice and reconsider the matter. The hazards of too much mistrust are equally hard to avoid but also difficult to conceptualize in those terms. It's natural to worry when mistrust goes fully neurotic, leaving you collapsed in a heap of self-obsession. But it can seem a puzzle to interpret either species of akrasia in terms of excess rather than deficiency. Isn't the failure to intend as you best judge or to act as you intend *a lapse*—perhaps, as Holton puts it, a failure of 'will-power'? Well, ask the same question of failing to trust your judgment or volitional dispositions at all: can we explain the neurotic's collapse as his merely lapsing? It seems not: accidie is not the same as exhaustion. The accidic agent feels no motivation to pursue what he values not because he's too tired but because he's in some way alienated from his evaluations. However we explain accidie, it seems appropriate to emphasize the active element in this alienation. And the same is true of the two species of akrasia: incontinence cannot be explained as the mere failure to be continent, and weakness of will cannot be explained as the will's failure to be strong.

If it seems paradoxical that a passive follow-through should express an agent's will, the reply is that not every form of passivity relinquishes control. An agent who chooses to act by trusting his judgment that he should φ does not relinquish control over his φing but does precisely what it takes to retain it. The work of the will here is to step aside and let action be guided by how you have made up your mind in judgment. This *re*-makes your mind insofar as (a) you are now guided by a counterfactual sensitivity to evidence of untrustworthiness in your judgment, triggering which would reopen practical deliberation, and (b) this sensitivity may misfire, rejecting judgment or intention as untrustworthy but *without* reopening deliberation. Akrasia would thus involve mistrustful activity—either rebellious choice or rebellious action—where rationality requires this species of trusting passivity. Akrasia is counter-*normative* agency because the sensitivity has misfired and has not led to redeliberation.

II

Before elaborating my argument for this position, let me make my view vivid with an analogy.

Imagine you're learning to dance in ballroom fashion, with your instructor as your partner. She leads, and you follow. Here are two interesting features of the case: *her* judgments and choices determine what happens as the two of you move across the dance floor, but what happens includes actions that *you* perform in just the sense in which you would perform them were you to lead. She determines the pace, the direction, even the precise shape of the physical movements you make, but this isn't like the case in which in an uncooperative moment you go limp like a rag-doll in her arms. The imagined actions are yours.[7] Nor is this like the case in which you learn by leading while she barks orders, or by imitating her movements, each of you with a separate partner. You perform the imagined actions, unlike these others, by letting your conduct be directly determined by the will of another. The two observations are related. It is your receptivity to the influence of the judgments and choices determining what you do that marks these doings as your actions.[8]

Now imagine you're an overeater learning to perform that delicate dance called *dieting*. The challenge, of course, is that you have nothing to guide you but your own judgment. It leads, and you follow—or try to. As in the ballroom, your moment-by-moment impulses are *just all wrong* and need to be guided by a force apart from them. Here too, rational agency is receptive agency. It is your receptivity to the influence of your own judgment that marks the doings as your actions.

The two cases are not perfectly analogous, of course. In the interpersonal, the influence on you is not merely your instructor's judgment but more pressingly her intentions or will. This suggests a different analogy with the intrapersonal: that your dieting cannot proceed unless you are suitably receptive to the influence not merely of your judgment *that* you should refrain from a second helping but of your choice or intention *to* refrain.

I'm interested in both analogies, since I'm interested in giving a univocal internalist explanation of the two species of akrasia.

III

I'll offer two broad arguments. First, my account captures both the logic and the phenomenology of akrasia. More generally, as I'll argue in the next section, it offers the best explanation of the normative guidance on which internalists rightly insist in the face of externalist postulation of a faculty of the will.

Consider this more detailed case. You've just started on your diet this week, and you're dining this evening at your in-laws. Your mother-in-law has concocted a fabulous cherry pie for dessert, whose potential she insists cries out for a scoop of ice cream. Did your diet-concluding or diet-intending self foresee this predicament? If not, why should you trust it now? If so, why should you trust a self that judged best or intended what, given your social awkwardness, will certainly amount to regrettable rudeness? Yet you know that there will always be such predicaments. The fundamental question is not how to avoid misweighing your best judgment as you form an intention, or how to maintain the strength of that intention, but whether to trust the self that drew that conclusion or formed that intention. Perhaps this was not the week to start your diet. Perhaps your diet needn't involve the renunciation of every dessert. Or perhaps it's time you simply forged ahead toward this goal and began awkwardly to learn the art of gracious refusal. There are of course these and other questions to deliberate, but a prior question is non-deliberative. It is a question of how to feel toward the earlier self whose judgment or choice created the predicament in the first place.

Note that phenomenology doesn't clearly distinguish the two kinds of akratic break. When you akratically took that second helping was it because you were weak in the face of temptation or because you formed a new intention out of accord with your best judgment? While there are clear enough cases on each side to make the distinction between incontinence and weakness important, there are many cases in the middle where either description is as good as the other. That is another reason, beyond the appeal of theoretical economy, to hope we can account for them in common terms. However we explain them, the two species of break should come out irrational for the same kind of reason.

The need for reasonable self-trust is logical: it follows from the assumptions that you can't act until you close and commit to deliberation and that acting is not only the causal but the rational upshot of deliberation. You'll be still deliberating, or redeliberating, insofar as you consciously wonder whether to trust your judging or your intending self.[9] And if you're deliberating or redeliberating your judgment or intention, you can't be acting on that judgment or intention. (Of course, you can conclude deliberation and commit to that conclusion without feeling certain that you're right: you need merely be as confident as your deliberative context requires.) As the rational upshot of a closed deliberation, action must be subject to rational guidance that is non-deliberative. As non-deliberative, this guidance must function in terms of a counterfactual sensitivity—that is, in terms of trust.[10]

But the account also captures key aspects of phenomenology. However conflicted you may still be about whether to φ,[11] your commitment to act acquires only as much guidance as a concluded deliberation can provide, and your action in turn acquires only as much as that commitment can provide. Both provisions pose questions for your self-relations: should you trust yourself?[12] If you don't trust your judgment or choice, there are two possible cases. In one, you reopen deliberation on the question whether or when to diet. In the other, you don't. Mistrusting your judgment or choice without reopening deliberation is irrational but not uncommon. That's what happens in each type of akrasia: you don't abandon your judgment that you should diet, or your intention *to* diet, but neither do you trust that judgment or intention. This is irrational because the attitudes of trust and mistrust are constitutively responsive to a rational norm. Trust is not the conclusion of a deliberation, but neither is it mere acquiescence. As I've argued, it is guided by a counterfactual sensitivity to evidence of untrustworthiness, triggering which would reopen deliberation. Mistrust thus rationally calls for deliberation. If you mistrust but do not reopen deliberation, your mistrust is pathological and therefore irrational.

In this respect trust functions like any emotion. Anger, fear, love, and compassion are pathological and therefore irrational if they do not, other things equal, lead you to act in certain ways in appropriate circumstances, where the emotions themselves determine criteria of appropriateness. If your reliance is trust, then you are necessarily irrational to the extent that you do not deliberate or redeliberate when confronted with evidence of untrustworthiness. As we've seen, you may nonetheless wind up relying on the person, or even on yourself, but that reliance will, unlike trust, be deliberated and cannot perform the role of trust in mediating the relation between deliberation and action. Where you mistrust, you must deliberate how to proceed.

It may seem that the second akratic break does not have this structure, since weakness of will requires merely an unreasonable revision of one's intention, not the deeper irrationality of rejecting it while failing to reconsider.[13] No doubt many cases in which the agent unreasonably redeliberates are worth chastising as 'weakness.' But the irrationality that puzzles is not just any unreasonable failure to follow through on an intention but a failure to follow through *without reconsidering* the matter. I'm arguing that we begin to understand such cases when we observe that the non-deliberative sensitivity that makes agency possible can be triggered without reopening deliberation.

Again, I don't claim that implementing this non-deliberative sensitivity will not require hard work. It may be very hard work indeed to stick to your judgment or resolution. I claim only that the task in question is best described as *maintaining appropriate receptivity* to your judgment or intention—that *that's* what's difficult in such a case. To see that there's no problem with this description of the difficulty, compare some cases from the dance floor. Say you're following the lead of someone with whom you're intensely angry. When the music stops you plan to punch this person in the face, but in the meantime you must maintain your composure—where that means, as you continue to dance, maintaining your receptivity to his lead. In another case, you're not angry with your partner but regard him as incompetent: when the music stops you plan to lecture him on how he should have led you through those steps. But in the meantime, even as you're writing that lecture in your head, you have to maintain receptivity to his lead—after all, you don't want him to get the idea that the incompetence lies with you! In each of these cases, maintaining receptivity to your partner's lead may prove enormously difficult.[14] And so it also may prove when the lead to which you must remain receptive is your own. It may indeed take effort to resist the urge to rebel against your own judgment or intention, but that's simply because there is no inconsistency in saying that it may take effort to remain appropriately receptive to this influence.

I don't deny that there are cases whose phenomenology suggests a role for 'strength of will.' I claim merely that the phenomenology in question should not be thus interpreted, given that the present interpretation is available. When your task is to 'stay focused' on your firm resolution to diet, your struggle is to remain receptive to that resolution. Your struggle in other cases may be less to remain receptive to a resolution than simply to keep from losing it. When your resolution to diet is less firm, you may keep vacillating over your judgment that today is the day to begin your diet. Assuming you don't abandon the judgment, your task is to resolve accordingly. What may present itself as 'strength of will' in retaining this resolution is actually rational-normative pressure to retain receptivity to judgment. I don't claim that phenomenology on its own reveals this normative structure. I claim merely that the structure can be discerned phenomenologically—if you know where to look.

IV

Now for the argument from explanation. As Gary Watson notes, externalist accounts of agency run the risk of conflating incontinence with existentialist 'radical choice:' "On the externalist view . . . going against reason must always come down to a choice among possible commitments."[15] Watson goes on to argue that we need to posit a faculty of the will to explain the possibility not of counter-normative agency but of normative uncertainty or indeterminacy, as for example when your reasons do not determine means to your end. But even here, he holds, your will is guided by judgment at least to the extent that "[w]hen intention fails to be guided by judgement, it fails to operate in its executive capacity—it fails to operate *as* a will."[16]

I agree with this last claim, but I don't think Watson or any other internalist has explained why it's true. Until we understand just what it is for judgment to guide or to fail to guide intention we won't understand what it is for an intention to operate or to fail to operate as a will. And if we don't understand *that*, then we won't understand akrasia. So until we have this explanation, the externalist is entitled to argue that only the postulation of an *externalist* faculty of the will, and of will-power, can explain the possibility of akrasia. We thus return to the two pressing questions for the internalist: How does judgment guide choice or intention? And how does choice or intention guide action? The question in each case is not *whether* it does but *how*.

My account has the virtue of providing these explanations. Rational guidance has two features: rational authority and motivational efficacy. A rational guide is simply a rational authority that as such motivates. On my view, judgment is rationally authoritative for choice (or intention), and choice (or intention) rationally authoritative for action, insofar as each is trustworthy in relevant ways, where such trustworthiness registers in the trusting not as a positive attribution but as the failure to trigger a counterfactual sensitivity to evidence that it should not be attributed. Again, it's just this that makes agency possible in the first place, since agency requires, in addition to the deliberative weighing of reasons, a rational receptivity to reasons taken as weighed—in other words, not only activity but this special sort of passivity. Much remains to be said about the specific nature of the authority in question. I don't claim that the authority *is* the species of self-trust just described, merely that it *registers* as such trust. In section IX, I'll suggest that the authority is aretaic and that a fundamental role for virtue thus lies in making agency possible. My present claim is only that internalists need to explain these species of non-deliberative rational authority, and that my account at least points in the right direction.

Its explanation of rational authority entails an explanation of how that rational authority can motivate. For trust already includes a motivational propensity: you trust someone—including yourself—only to the extent that you are disposed to let the person influence your motives. So if it is something internal to trusting receptivity that explains how your judgment authorizes your choice or intention, and again how your choice or intention authorizes your action, we need add nothing to get an account of motivational efficacy.

The account explains akrasia as what happens when you mistrust the source of rational authority over your motives without reconstituting that authority through further deliberation. Your motives thus come unglued from what authorizes them, but in a way that is intelligible—since you would not have been irrational had you taken that extra step and reopened your mind. You judge that you shouldn't have dessert tonight but fail to choose accordingly. Or you intend to pass on dessert but fail to act accordingly. We can imagine a variant on each case that would not be akratic: it occurs to you that your sugar cravings are unexpectedly interfering with your concentration, and this evidence of your untrustworthiness in judging or intending (since you didn't foresee the problem) leads you to wonder if tonight is really a good time to begin your diet. That is, you come to mistrust your judgment or intention, and the mistrust points a path forward, practically speaking, by reopening deliberation. In the akratic version of the case, by contrast, though you mistrust the judgment or intention your mistrust does not reopen deliberation. The mistrust deauthorizes the judgment or intention—in the sense that you do not treat it as authoritative—but you do not reconstitute that authority by redeliberating and reaching the conclusion that you should not start your diet tonight. It is intelligible that you should fall into this species of irrationality simply because changing your mind has two distinguishable steps. Though they frequently happen simultaneously, we can distinguish the step of mistrusting your judgment or intention from your first step toward a different deliberative conclusion.

Sometimes, after all, you mistrust your judgment or intention without knowing how to go about approaching the question differently. Sometimes, that is, you continue to judge that you should φ or continue to intend to φ despite the fact that you mistrust that judgment or intention. It 'feels' wrong to begin your diet tonight, for example, so you don't—even while continuing to judge that you should or intending to. Your intelligible mistake is to let your mistrust take you only halfway toward a change of mind. It's as if the mistrusted judgment or intention were someone else's, and you could fulfill your agential responsibilities simply by resisting it. 'I don't trust that influence,' you seem to be saying—without realizing that the influence is your own.

The normativity of trust relations doesn't in general preclude such a refusal of trust. Simply refusing another's invitation to trust without deliberating what to do instead can be perfectly reasonable. ('Hey, get off that couch and let me give you a leg up to this bird's nest.' 'Um, no.') But when the one inviting your trust is you yourself, you rationally can't just refuse. You can rationally refuse only by reconstituting the authority by appeal to which you issued the invitation. The difference does not lie in any deep self-other asymmetry but simply in the fact that the invitation in the second case is your own. You have to be responsive to it because you *are* responsive to it insofar as it is *yours*. We aren't talking about a case in which the judgment or intention operates subconsciously, or in which you're confused about your identity. From this angle, your mistake in akrasia would lie in a form of self-deception about your identity: you pretend to yourself that your own judgment or intention can be rejected as if it were that of another. You don't like the fact that you have committed yourself to dieting this evening, so you pretend that the commitment is a coercive intervention from a perspective not your own.[17] I don't claim that the pretence need be in play. This is merely a way of making the pathological trust relation vivid.

We can likewise explain the possibility of choosing under judgmental uncertainty or indeterminacy in terms of what is not only actually the case but a truism: that trusting your judgment takes you only so far. You judged that you should grab a box of Oat Flakes; you didn't judge which box. You judged that you should knock the robber to the ground; you didn't judge precisely how. Still, when just that box winds up in your cart and the robber winds up on the ground nursing just that bruise, it was because of your choice. We can agree that judgment does not—indeed, cannot—authorize every aspect of what you do when you act while nonetheless insisting on an internal link between judgment and action. We thereby explain how the exercise of judgment makes choice or intention function as your will. Your will is what determines your actions as such, not what determines your every movement when you act. (In like manner, choosing the box on the left or to grab the robber's right leg doesn't resolve every question of muscular contraction.) Insofar as choice or intention fails to be guided by the authorization of judgment, either because you choose or intend counter-normatively or because of normative uncertainty or indeterminacy, it isn't functioning as a will. Counter-normative agency is thus an explicable by-product of agency, while choosing where judgment gives out merely reveals that agency unfolds at a level that is less fine-grained than the levels at which we can be described as either doing things or choosing to do things (just as the level of agency is much less fine-grained than the level at which we can be described as contracting our muscles).

My thesis, then, is twofold: your choice or intention functions as your will when it expresses an executive authority that derives from its trusting receptivity the verdict of your judgment, and your behavior manifests your will when it manifests a trusting receptivity to the influence of your choice or intention. Choice or intention registers thus receptively the authority of trustworthy judgment, making it motivationally efficacious. My main argument is that this thesis best explains the possibility of the two species of akratic break.

V

Let me turn to a trio of objections to this argument. First, it may seem wrong to say that you mistrust your judging self when you fail to choose or intend as you judge best, since you needn't deem your judging self untrustworthy. Second, it may seem that I have not really explained either species of akrasia, since I have not explained the difference between irrationally mistrusting your judgment (whether in willing or in acting) and merely being prevented by some arational force from trusting your judgment. And third, it may seem that my account ignores present-directed intentions and generally imposes a temporal ordering among the elements of rational agency that doesn't define them. Can't you simultaneously judge that you ought to φ, choose to φ and φ? The acts needn't, it seems, unfold in a temporal sequence. I'll treat the first and second objections in this section. The third will require a recodification and elaboration of my position that I'll pursue in the following four sections. The recodification will in turn provide a framework for my account of the aretaic basis of practical-judgmental authority over the will.[18]

Consider this case. You deliberate and reach the judgment that you ought to get right to work this morning. Then without reconsidering the matter you choose to linger over the newspaper, intending to keep reading merely, as you put it, "till I finish this cup of coffee." Perhaps you do form the intention to get to work but 'find yourself' nonetheless giving in to the temptation to linger. The coffee is of course almost cold, and you know you have no disposition to finish it. But you continue to judge, all things considered, that you ought to work this morning.

On my account, your mistake lies in the fact that you fail to trust your judgment but without doing what follows from that mistrust, namely reopening deliberation on what to do this morning. Of course, one reason why you don't reopen that deliberation is that you don't believe you made a mistake when you conducted it earlier: you continue to judge, all things considered, that you should work this morning. Given this judgment, how could failure to reconsider be a mistake? The problem is that you're failing to let yourself be governed by that judgment. That is, you're failing to trust it. Since it's your judgment, you're therefore failing to trust yourself. You don't, of course, *believe* yourself untrustworthy. In fact, you believe yourself trustworthy. But that belief is irrelevant to whether you trust.

Compare an interpersonal case. 'Go on up,' your climbing instructor assures you. 'I'll use this rope to keep you from falling if you lose your grip on the rock.' You trust him, deeming him highly competent, so off and up you go. Five feet up, however, you stop. You can't bring yourself to go further. It seems you don't trust him after all—but without deeming him untrustworthy. You can't imagine a more competent climbing partner and wouldn't know how to go about redeliberating whether he's worthy of your trust. Still, you can't bring yourself to trust him. Your judgment simply doesn't govern your will.

This sort of thing happens so often that we rightly don't find it at all paradoxical. Trusting another requires more than the judgment that he or she is worthy of your trust. Yet we puzzle over the intrapersonal analogue: how could you fail to trust your judgment when you don't deem your judgment untrustworthy?

My proposal is that we model the intrapersonal on the interpersonal in two steps. First, note that you might judge and even feel perfectly confident that your climbing skills are up to the route you've mapped yet—perhaps spooked by a recent fall—find yourself unable to trust them. Here you fail to trust yourself in exactly the way that you failed to trust your climbing instructor. Second, make the competence not a physical competence figuring in how you will execute the action but a mental competence figuring in how you deliberated whether to perform it in the first place. The parallel nonetheless holds. Just as you can mistrust your own physical capacity without deeming yourself unworthy of that trust, so you can mistrust your own judgmental capacity without deeming yourself unworthy of that trust. After all, you could as well mistrust your instructor's judgment—say, on whether this route is within your competence, or simply on how distracted by conversation to let himself get while spotting you—as his physical capacity to keep you from falling.

A pithier way to make the point is to note that if you mistrust your judgment on some question it isn't surprising that you might at the same time *judge* that you are trustworthy on the question. The latter is simply part of what you mistrust. To mistrust your judgment in a given respect entails mistrusting any judgment that you might make that you are trustworthy in this respect. To flip it around, if you are not receptive to your own judgment on the question whether you should φ, you can't expect to make yourself so by reflecting that you *judge* that you should be receptive. This is what we should expect if the problem is inappropriate receptivity: you can't make yourself appropriately receptive by trying harder. The problem is that you're trying too hard—manifesting discounted worries as anxiety and generally second-guessing yourself—and you need to find a way to let this activity subside. You need to relax and let your own judgment guide you—either this one or, after redeliberation, another.

It is thus crucial to my account of receptive agency that mistrusting someone is not the same as believing him or her untrustworthy. When I speak of trust and mistrust, I do not mean the attitudes of deeming-trustworthy and deeming-untrustworthy. Those doxastic judgments may or may not inform trust and mistrust, whose criteria of appropriateness, like those of other emotions, govern the subject without figuring directly in the subject's judgmental attitudes. Just as love presents its object as lovable without entailing any judgment of lovability, so trust presents its object as trustworthy without entailing the judgment that she is. Just as you can without doxastic inconsistency love or hate those whom you do not believe worthy of the emotion, so can you trust or mistrust where you do not believe the object deserves it.

What then of cases in which you fail to be guided by your own judgment only because you are gripped instead by an arational force such as a compulsion? What if when you give in to the temptation to linger over your newspaper, what you 'find yourself' doing is not reading but obsessively rehearsing an embarrassing memory of which an article has reminded you? You'd like nothing more than to relax and let yourself be guided by your judgment that you need to get to work, but as long as this force has you in its grip you simply cannot. Is it correct in this scenario to say that you 'mistrust' your judgment? It seems not: *you* are not doing much of anything in an agential sense except trying to resist this force that is acting upon you. (I assume that the obsessive memory is not merely masking an irrational mistrust in your judgment, providing you with a convenient excuse: "Hey, I'm not lingering but in the grip of a compulsion here!" I assume that your obsessive memory is more like a compulsion to count the 'the's on the page than like daydreaming.) So if you fail to intend as you judge or to act as you intend in *this* way, it does not count as the philosophically perplexing sort of akrasia—that is, as the sort I am trying to explain. (Or we could reserve ‘akrasia’ for the perplexing sort of break and say that this is not akrasia at all.) There is nothing philosophically perplexing along these lines about compulsive action or ideation. The mental activity or behavior in question simply does not qualify as choice, intention, or action.

Still, do I not owe an explanation of the distinction on which I am relying, between a compulsive or otherwise arational failure to trust and the irrational mistrust distinctive of akrasia? Well, I have just given it. At least, I have explained my entitlement to rely on the distinction. We begin with a puzzle over how it is possible for you to choose or intend contrary to your own best judgment, or to act contrary to a choice or intention informed by that judgment. A grasp of that puzzle presupposes a grasp of the distinction in question, since the puzzle presupposes that you are not merely gripped by a compulsion or some other arational force. Noting this feature of the puzzle does to some extent explain the distinction, since it articulates a dimension of the arational/irrational distinction in play here. It is not, of course, a full explanation of that distinction. But in the dialectical context at hand it is all the explanation I owe.

As a variant on the second objection, one might worry that my account renders all akrasia merely arational, since I have not explained how akratic choice, intention or action could qualify as such (as opposed to something like compulsion), given that it fails to be informed by actual judgment. (As we’ll see in section VII, it is informed by an ‘as if’ judgment: you do act as if your best judgment supports you.) Here again I need merely clarify my explanandum. I regard the phenomenon of akrasia as most fundamentally manifested in the akratic *break*, which I explain as an irrational failure to trust your judgment. This failure is most strikingly present when the akrasia occurs by omission, which is why I’m now focusing on a case in which the akrasia is most naturally described as a deficit: you judge that you ought to do something (get to work) other than what you’re doing (reading the paper), and the question is how you could fail to treat that judgment as a rational guide. In such a case we don’t need to explain how you could continue to read the paper, since that’s what you were doing anyway. What we need to explain is how you could fail to at least try to get to work. To the worry that I haven’t explained how the akratic pie-eating in my earlier example could count as anything but a compulsive spasm, I reply that I am not here giving an account of attributable agency in general.[19] That akratic pie-eating falls short of a paradigmatic instance of attributable agency does not entail that it falls short of such agency altogether. An account of akrasia should explain how each species of akratic *break* is possible. It needn’t add up to a theory of attributable agency.

VI

My account needs elaboration in a different respect. My approach to akrasia assumes that a practical judgment can possess a distinctive species of rational authority, but I haven't yet explained the nature or basis of that authority. I'll now locate the authority in the species of trustworthiness that you presume when you make a practical judgment.

I'll begin by framing that explanation within a more general issue. It is often assumed that a practical judgment to φ at t is the judgment that you have reasons (perhaps, conclusive reasons) to φ at t. That is, it is often assumed that a practical judgment is a doxastic judgment (perhaps, a belief) about your practical reasons.[20] There are two problems with that assumption. It obscures some important differences between doxastic and practical judgment. And it overlooks the important difference between a judgment *about* reasons and a judgment *informed by* reasons. Let me explain each problem in turn, beginning with the second.

Compare my account with an account of the rational relations among judgment, choice or intention, and action recently offered by T. M. Scanlon. Scanlon maintains that deliberation issues in a judgment, not (simply) that you ought to φ, but that you have conclusive reason to φ.[21] Scanlon holds that, while choice or intention needn't be informed by a judgment about reasons, it is nonetheless the role of judgment to provide such higher-order guidance where appropriate. On my view, by contrast, deliberation issues in a first-order judgment informed by reasons, not in a higher-order judgment about reasons. (Of course, you may also make such a higher-order judgment. But you needn't.) It is just this that allows me to argue that there is judgment whenever there is choice or intention, since each of the latter embodies a commitment to the deliverance of judgment. The fact that the guidance provided by judgment is first-order is what makes possible this species of commitment: you judge that you ought to φ, a judgment that becomes a choice or intention to φ as soon as it is received in self-trust. On my view, commitment to the course of action is mediated by the primary commitment to the deliverance of judgment. Your judgment is what you trust, what holds your allegiance, and only thereby do you commit yourself to performing that action. On Scanlon's view, by contrast, the commitment in a choice or intention cannot be *to* the deliverance of judgment—since the contents differ. So it must be to the course of action directly.

My account therefore generates an internalist approach to agency unavailable to Scanlon. I treat the action as manifesting a rational responsiveness not merely to the executive authority of choice or intention but to the rational-deliberative authority of judgment. Even apart from what I am arguing are the attractions of an internalist approach, it is a problem that Scanlon has ruled out such internalism by definition. If the content of a practical judgment is higher-order, then by definition it cannot provide the rational guidance claimed by internalism. Even if internalism is wrong, it should not be ruled out by the stipulation that the guidance provided by judgment is higher-order.

Nor should practical judgment itself be defined as doxastic. A view of judgment should do justice to the status of a judgment as the conclusion of a deliberation. Since there are two importantly different species of deliberation, doxastic and practical, there are two importantly different species of judgment. Doxastic deliberation addresses a question of the form *whether p*. Practical deliberation addresses a question of the form *whether to* φ. We should therefore not construe practical deliberation as addressing a doxastic question: *whether there is conclusive reason to* φ.

One problem with defining practical judgment as doxastic is that practical reasons arguably cannot register as such within a doxastic deliberation. If evidentialism is true, as is widely believed, then only evidential considerations can register within a doxastic deliberation. Assuming evidentialism, the deliberation *whether there is conclusive reason to* φ must be guided by evidence of the existence of such reasons. But the deliberative weight of a practical reason is not purely evidential. Or rather, it is purely evidential only on the very controversial view that there is no necessary connection between possessing or weighing a reason and feeling motivated to act on it.[22] Unless we simply assume this controversial view, we cannot assume that weighing evidence that you have a practical reason is the same as feeling the deliberative force of that reason directly. Of course, one could try to argue that evidence of a practical reason, unlike other evidence, *is* intrinsically motivating. But such an argument seems fundamentally misconceived. The best way to debate the similarities and differences between doxastic and practical normativity does not begin from the assumption that what is distinctive of practical normativity must emerge in a deliberative framework already defined as doxastic.

VII

My own framework locates the difference between doxastic and practical normativity in the content of the self-trust that deliberation aims to secure. It therefore locates the difference in the nature of the trustworthiness presumed by the judging self. I lack space to consider parallel issues concerning doxastic judgment.[23] The details I'm about to present supplement but are not required by my argument through section V.[24] I'll approach these structural issues by way of the third objection described at the beginning of that section.

I've been talking as if practical judgment leads to choice or intention, and choice or intention leads to action. But what of cases in which the agent's judgment and choice or intention lies *in* the action? On the tennis court, for example, you may choose to return a fast serve with a passing shot, since your opponent is charging the net. Here judgment, choice and action seem to happen simultaneously: you barely have time to get your racket back, much less to contemplate where to send the ball. Still, you do deliberate, form a judgment, choose on the basis of that judgment, and act on the basis of that choice. What reveals that you do all these things is that we can imagine each species of akrasia in such a case: perhaps you are unmoved by your own judgment about where to place the shot, or perhaps by contrast you are irresolute about the very practical commitment that you then and there form. Since, as I noted in section III, it can be hard to imagine a non-philosophical point in distinguishing such cases, it's easy to feel unconfident how to categorize a given instance of akrasia. Diagnosis is inevitably retrospective and uncertain, but we can distinguish two sorts of case. Assuming that the problem is not mere bad aim—that is, garden-variety incompetence at getting body and world to cooperate in executing your choice—'But I *knew* I needed to pass him!' suggests a failure to let judgment govern choice, whereas 'But I *meant* to pass!' suggests a failure to let choice govern action. (In either case, commentators will rue that you're 'thinking too much.') How then could judgment, choice and action figure as aspects of a single occurrence?

One could, I suppose, hold that judgment, choice and action are never simultaneous and simply insist that the sequence in which you make, commit to and act on a judgment may unfold very quickly. If one takes that route, then the present issue gives way to a different and I think less tractable issue: how do you pull off these transitions so quickly? You'll have to count as deliberating for part of that eye-blink, judging for another part, committing yourself to the judgment for yet another, and only then performing the action. The model seems pure fantasy. Rational agency is neither like producing an action on an assembly line nor like rolling a cigarette or firing from a holster with what appears to be a flick of the wrist. Of course, it can unfold in stages: deliberation, judgment, choice, intention, action. But it can also happen all at once. And even when it doesn't happen all at once, we can view judgment, choice, intention, and action as aspects of that unfolding event, not parts.

I'll now elaborate my theory in such non-temporal terms, showing how judgment and choice or intention are aspects of the disposition to act, not temporal parts of it. At least, they are aspects of the disposition when judgment governs action. My aim is twofold: to codify my view of rational agency as receptive agency, and to show how such reception is not a temporal process. My account achieves that aim by reductively analyzing the acts and states of judgment, choice and intention in terms of complexly self-referential dispositions. Each act or state is therefore nothing but a codification of the dispositions figuring in rational action itself, whether or not those dispositions come to exist in a temporal sequence.[25] As we'll see, the challenge is to articulate precisely how the rational authority of practical judgment registers as a species of trustworthiness.

Since it's simpler, I'll start with my account of choice, which I'll reformulate as an account of intention later in this section. Call being disposed to act as if p 'accepting' that p.[26] By 'act' I include mental acts as well as physical acts: the disposition in question may in some cases be restricted to thinking or feeling as if p. Accepting that p is what imagining that p, supposing that p, pretending that p, and actually judging that p all have in common in terms of the subject's behavioral dispositions. I don't claim that there is a distinct kind of act—something one could actually do—that amounts to *merely* acting as if p in this sense. 'Act as if you're under snowball attack,' cannot be fully implemented apart from a context indicating the presumed point of the behavioral routine, which will clarify whether you're to, for example, imagine the projectiles or actually expect their arrival. All that's important is that accepting that p not entail judging that p. As long as we understand what this behavioral core shared among these attitudes would roughly involve—conceiving it not as a self-standing act but by abstraction from fuller acts—we understand enough for my account to proceed.

So then:

> (C) S chooses to φ at t[27] iff S accepts: that she judges (all things considered) that she ought to φ at t and in so judging is practically wise.

I'll say more about the role of practical wisdom in choice in section IX. For now, note that, on (C), choosing to φ directly includes accepting that—i.e. being disposed to act as if—you judge that you ought to φ.

This account of choice may seem to make incontinence impossible, since it may seem to make it impossible to choose against your judgment. But that is not what (C) says. (C) doesn't say that you cannot choose to φ at t without judging that you ought to φ at t. It says that you cannot choose to φ at t without *accepting* that you judge that you ought to φ at t. And the latter claim does ring true. When you choose to eat the slice of pie that you judge you should not eat, you do accept that—that is, act as if—you judge (all things considered) that you should eat that pie. Of course, you don't actually make that judgment; but the point is that you act as if you do. As I put it in section IV, when you choose against your own best judgment you act as if that judgment is not yours. We can now put it this way: you act as if you've instead judged in accordance with your choosing (though you have not actually made that judgment). You act not only as if you've judged that you should eat that slice of pie but as if so judging would constitute you as practically wise. This is how you stake a practical commitment—whether in accordance with your actual judgment or not.

You can, of course, choose to φ at t without having the *belief* that you're practically wise. Again, to say that you 'accept' that you are practically wise is not, as I'm using the term, to say that you've formed any judgment or belief. It's to say that you are disposed to act as if you are practically wise—a non- or pre-doxastic disposition that is compatible with your pretending or merely supposing that you are practically wise. There is, as we'll see, a degree of self-consciousness in such a disposition, but it does not (yet) involve the self-consciousness characteristic of either practical or doxastic judgment.[28] Moreover, accepting that p entails no species of commitment to *p*—neither doxastic nor practical.

My account of practical judgment will be reductive: it will reduce judging that you ought to φ to a complex of dispositions, none of which on its own possesses the rational authority characteristic of practical judgment. The practical judgment that you ought to φ neither is nor entails, of course, the simple disposition to φ—since that would rule out the akratic failure to be disposed to act as you judge best. Nor is it the mere second-order disposition to treat yourself as disposed to φ—since that would omit the rational dimension of the self-relation. It does, however, build on that second-order disposition. Practical judgment adds to the second-order disposition a further disposition to treat the fact that you have that second-order disposition as giving you a reason to realize it in first-order action. And not just any reason: you're disposed to treat the second-order disposition as giving you a preemptive reason to act, which I define as a reason to act without further deliberation. (I'll say more about such reasons presently.[29]) In sum: a practical judgment is the disposition to treat the fact that you are disposed to treat yourself as disposed to φ as a preemptive reason to φ. Why would you have this complex disposition? Because, as we'll see, you've closed deliberation by embracing the presumption that—by forming the disposition to act as if—your trustworthiness on the matter gives you the relevant species of rational authority.

Here is a formulation of that account along quasi-Gricean lines:

(PJ) S judges (all things considered) that she ought to φ at t iff S accepts:
  (i) that she has a preemptive reason to φ at t,
  (ii) that she accepts (i), and
  (iii) that (i) because (ii).

For ease of formulation I'll henceforth omit the 'all things considered' clause, unless I'm contrasting practical judgments proper with mere *pro tanto* judgments, which lack the species of rational authority of characteristic of the former. It should, however, always be understood.

According to (PJ), then, judging that you ought to φ at t is accepting that you have a preemptive reason to φ at t simply through your accepting that you have that reason. But the explicans of (PJ) does not entail the explicans of (C). Accepting that you have a preemptive reason to φ at t, and that you accept this, is compatible with not choosing to φ. You manifest the first acceptances by, for example, being disposed to reply 'φ' to the question 'What should you do at t?' and, most fundamentally, by no longer being disposed to deliberate the question what to do at t—that is, by being disposed to treat the matter as closed, no longer considering the alternatives to φing at t. You can coherently have these dispositions while lacking both the disposition to φ at t and the disposition to treat yourself as practically wise in judging that you ought to φ at t.

As a self-doubting dieter you may treat the question of what to eat as closed without being disposed to eat what you judge best. And the explanation why you are not disposed to eat what you judge best may be that you are not disposed to treat yourself as practically wise in making that judgment. That is, you accept that your acceptances give you a preemptive reason not to eat that pie, but you don't accept that any of those acceptances constitutes you as a real executive authority over your conduct. I'll say more about the role of practical wisdom in constituting executive authority in section IX. For now, let's make do with the rough intuitive notion at work in the thought that when you conclude practical deliberation with a judgment you treat yourself as—to that extent, at least—an authority on whether you should φ. If you don't view yourself as an authority on whether you should φ but are still faced with the question whether to φ, then you ought to stop and redeliberate. The irrationality of akrasia lies in the fact that your choice or action shows that don't treat yourself as an authority on what to do, yet you didn't stop and redeliberate. For now, let's think of practical wisdom simply as the personal attribute informing such authority.

As a self-doubting dieter, then, you treat yourself as a rational authority on what you should do: you are disposed to treat the question of whether to eat the pie as settled by your deliberation and not at all disposed to redeliberate. But at the same time you do not treat yourself as an executive authority on what to do: you ignore or defy your judgment, willing or acting contrary to its dictate. But what could constitutive you as an executive authority on what to do other than your status as a rational authority on what you should do? The tension manifested by such a failure to reopen deliberation is what defines incontinence. And weakness manifests a similar tension: by not acting on your choice or intention, but without reopening deliberation, you reveal that you do not treat your choice or intention as possessing the executive authority that you're claiming for it. The trick is to see how practical judgment leaves room for these self-relations. According to (PJ) it does so by leaving it open whether you actually implement the authority claimed by your judgment. Though you of course *may* (and typically do) both choose to φ at t and φ at t when you judge that you ought to φ at t, you needn't do either. Sometimes you judge that you ought to φ at t without choosing to φ at t; sometimes you judge that you ought to φ at t, choose to φ at t but fail to (even try to) φ at t, and sometimes—thankfully—you do all three.

That formulation suggests, of course, that it's merely incidental to judgment that it governs choice and action—the very thesis that my internalist approach to agency purports to prove false. How, then, does judgment aim at such governance? By (PJ), it does so by presenting itself—that is, the judging self—as worthy of the acting self's non-deliberated (that is, not further deliberated) trust. According to (PJ), when you judge that you ought to φ at t, you presume that you thereby come into possession of a reason to φ at t without redeliberating the matter—a sort of reason that I'm calling a *preemptive* reason to φ at t. This is not a reason that could have figured in any deliberation on whose basis you came to judge that you ought to φ at t. Rather, it's a reason whose authority derives precisely from the trustworthiness of that judging self. Moreover, it's a reason to do just one thing: act in accordance with that judgment. So this is not a reason about which one might have bootstrapping worries. You don't possess the reason unless you really are trustworthy in judging that you ought to φ at t, and you can't simply bootstrap your way into that status.[30] I'll say more about this trustworthiness—a species of practical wisdom—in section IX. My present point is merely that when you form a practical judgment you presume that you have that status, since if you didn't you wouldn't feel entitled to close the deliberation.

We can see how the presumption of authority works by codifying it within the forward-looking perspective of intention. Let $S_{F/A}$ be a subject separated into F, the earlier self that forms and then continues to have an intention, and A, the later self that acts on it. Now:

> (I) $S_{F/A}$ intends to φ at t iff F expects and desires that A will at t have a preemptive reason to φ at that time simply through A's memory of that desire.[31]

By a 'preemptive' reason, again, I mean a reason that does not engage your motives through deliberative (or redeliberative) reflection. Here, the preemptive reason would be a reason to follow through on your intention without redeliberating whether it was well formed. By (I), an intention is just the conjoining of an expectation and a desire, where both the expectation and the desire pertain to the agent's possession of a specific preemptive reason: a reason the agent counts as having if and only if he remembers the desire that he should have it. How could you come to have such a reason? Again, as I'll explain further in section IX, to possess such a reason you must be exercising the virtue proper to practical judgment, which I follow tradition in calling practical wisdom. As (C) makes clear, to commit yourself to a course of action you accept that your judgment that you ought to pursue it constitutes you as practically wise. What we see in (I) is simply how that commitment expresses the presumption that you'll be acting rationally—that is, from a reason[32]—if you follow through on it without further deliberation.

(I) formulates that thought using a doxastic-judgmental verb, 'expects,' rather than the pre-judgmental verb, 'accepts,' that figures in (PJ). So (I) is not, and cannot on its own be, as ambitiously reductionist as (PJ): intention requires a doxastic judgment, not merely dispositions. When you follow through on an intention you may no longer be otherwise disposed to perform the act. That's one reason why we need intentions: to get ourselves to follow through on practical commitments just as such, without further motivational prompting. We can nonetheless explain the normativity of that 'expects' in terms of (PJ) and (C), construing it as the forward-looking aspect of (PJ)'s and (C)'s complex acceptances. (PJ) depicts you as accepting that you have a preemptive reason to φ at t simply through accepting that you do. (C) adds to that the acceptance that this complex acceptance constitutes you as practically wise. (Again, I'll say more about practical wisdom in section IX.) And (I) codifies that commitment to φing at t diachronically, from the perspective of the self who at t will do the φing. (I) adds to (C) the claim that in a diachronic case you must expect—that is, make a doxastic judgment—that you'll appear trustworthy to your future self.[33]

To see how this works, consider two cases. First, imagine that the time of action, t, is close enough at hand that the complex acceptance in (C)—that is, that disposition—can be relied on to carry you through the action: you'll φ at t because—motivationally—you'll continue to accept that the acceptance in (PJ) constitutes you as practically wise. Second, imagine that t is far enough into the future that you cannot rely on this disposition to carry you through the action. Here you'll need to create a new disposition, one that doxastically thematizes your own status as the source of a preemptive reason to φ at t. This thematic emphasis is not present in the dispositions cited in (PJ) and (C). There your self-consciousness need not involve doxastic judgments: you accept that your acceptances constitute you as practically wise, but you needn't form the judgment that they do. A judgment about your own status as the source of a preemptive reason becomes necessary only when it becomes necessary to form an intention, and for the simple reason that only such a judgment can get you to φ at t without redeliberating the matter. You have to expect—that is, judge—that you'll remember your desire to φ at t. The motivation that carries you through is the desire that, by (I), you expect you'll remember you have. We could say that your memory that you had this desire gives you a new desire to φ now (i.e. now that it's t), but I think it's more plausible to say that the memory of this desire simply reactivates it. 'That's right, now is when I want to φ,' you might say to yourself as you remember the desire. You've had the desire all along, since you've all along intended to φ at t, but only this memory could put you in motivational touch with the desire. And that's exactly what you expect and desire when you form the intention: not only that you'll be motivated by the desire but that this motivation will express your status as a source of preemptive reasons. Even when the self-relation stretches over years or decades—imagine, say, booking a flight to your fiftieth high-school reunion simply because you've 'always' intended to go—it manifests a passive receptivity to trustworthy practical judgment.

VIII

Each of the accounts can be stated in a sentence—as (PJ), (C) and (I)—but much remains to be said about how they work. The first explanatory challenge is simple: How could you have a preemptive reason for an action *because* you accept that you do? A way to express (PJ) is to say that judging is inviting reason-generating self-trust. But how could self-trust generate a practical reason? In this section I'll characterize the self-relation in more detail; in the next I'll say how it can generate a practical reason.

Consider what (PJ) says. You judge that you ought to φ, it says, not merely by accepting that—i.e. being disposed to act as if—you have a preemptive reason to φ. You're disposed to act as if you have such a reason when pretending that you ought to φ, but when you pretend that you ought to φ you don't (except rarely) judge that you ought to φ. A distinctive species of *commitment* to the course of action figures in judging that you ought to φ that does not figure in imagining, supposing or pretending that you ought to φ. These other acts embody other species of commitment, but not the one distinctive of practical judgment. Nor should we conflate the commitment distinctive of practical judgment with the commitment distinctive of choice or intention: the possibility of incontinence shows that the former does not entail the latter. How, then, does a practical judgment commit the judger?

It doesn't help to say that your judgment manifests commitment to the *truth* of the proposition that you ought to φ. Supposition and pretense embody their own species of commitment to the truth of that proposition. When you suppose or pretend that p, you suppose or pretend that *p* is true, and in that respect you're—imaginatively, we might say—committed to the truth of *p*. It doesn't help, therefore, to note that judging, whether practical or doxastic, aims at the truth—or, if you've got metaethical scruples about the notion of practical truth, that practical judging aims to *get right* the matter you're deliberating. So, in its distinctive way, does every other cognitive propositional attitude and act.[34] The question is *how* judging aims to get it right.

To this question (PJ) provides an answer. You aim to get practical judgment right, it says, by aiming to φ only when your acceptance that you have a preemptive reason to φ gives you a preemptive reason to φ. Consider in turn the two clauses of this necessary condition, (ii) and (iii) in (PJ). To say, with clause (ii), that you accept that you accept that you have a preemptive reason to φ is to say that you're disposed to act not only as if that's true (i.e. that you have the reason) but as if you're disposed to act as if it's true. This requires merely a minimal degree of self-consciousness—not that you think about yourself or your dispositions but that you're disposed to act as if you have those dispositions. (And remember: the possibility of incontinence shows that being disposed to act as if being disposed to φ gives you a preemptive reason to φ needn't involve the disposition to φ. Perhaps you don't trust the first disposition.) Clause (ii) describes a self-consciousness involved in the reflexivity of human thinking, which needn't yet involve self-awareness.[35]

The reflexivity of judging does not yet distinguish judging that you ought to φ from imagining, supposing or pretending that you ought to φ. None of these acts is a mere disposition to act as if you have a preemptive reason to φ. Sometimes we *are* merely disposed to act as if we ought to φ. If stricken with an obsessive-compulsive disorder you may be disposed to act as if your hands are always dirty, washing them repeatedly. But that's not yet to judge that you ought to wash your hands. It's not even to imagine, suppose, or pretend that you ought to do so. Of course, it's possible that as a result of these dispositions you do come to imagine, to suppose, or even (by actually deliberating the question) to judge that you ought to wash your hands. But the ailment itself is merely that you are stricken with uneasy thoughts of uncleanliness that lead you to bouts of behavior. Though it may cause you to imagine or even to judge that you ought to wash your hands, the ailment itself, a mere disposition, does not include the reflexive dispositions involved in these attitudes. By contrast, each of these cognitive acts and attitudes—imagining, supposing, pretending, as well as judging—entails that you are disposed to act as if you ought to φ, insofar as you are, in part because you're (in the right way) disposed to act as if you're disposed to act as if you ought to φ.

Clause (ii) involves more, of course, than mere reflexivity. Call a disposition to act as if you're disposed to φ a *reflexive disposition* to φ. Clause (ii) does not assert that you have a reflexive disposition to φ; it asserts that you have a reflexive disposition to act as if you have a preemptive reason to φ. But that more complex reflexive disposition also fails to distinguish judging that you ought to φ from other attitudes. When you imagine that you ought to φ, you're imaginatively disposed to act as if you have a preemptive reason to φ. (For example, you imagine yourself entitled to act on the judgment without redeliberating, which is to imagine yourself possessed of the relevant species of rational authority.) When you suppose that you ought to φ, you're suppositionally disposed to act as if you have a preemptive reason to φ. And so on. Again, each cognitive attitude embodies its distinctive species of commitment to φing. The commitment is merely the distinctive way in which you're reflexively disposed to act as if you have a preemptive reason to φ.

The differentia of judging lies instead in clause (iii), which says that to count as judging that you ought to φ you must be disposed to act as if your disposition to act as if you have a preemptive reason to φ gives you a preemptive reason to φ. No other act or attitude embodies that implication. When you merely imagine or suppose that you ought to φ, you don't presume that your dispositions actually *give* you a preemptive reason to φ; you merely manifest a disposition to act as if you possess it. But how could an act or attitude satisfy this differentia? Here we get to the heart of the matter. How could your reflexive disposition to act as if you have a preemptive reason to φ give you a preemptive reason to φ?

If it's this reflexive disposition to act as if you have a preemptive reason to φ that gives you a preemptive reason to φ, then it's a fact about you that gives you a preemptive reason to φ. And if in judging that you ought to φ you represent yourself as authorized to φ by this fact about you, then in judging you represent yourself as authorized by a species of self-reliance. As I'll explain further in the next section, this is not the insane thesis that you bootstrapped your way into possession of the reason. It's merely the tautology that when judging you can't avoid relying on your own status as reliable in judging.

Since parallel tautologies are equally true of the other acts, we must say specifically what reliability in practical judging comes to. We can again contrast judging with imagining, and it helps to begin with an interpersonal case of each. Say you're playing a game of make-believe with some children, in which slabs of mud figure imaginatively as delicious pies. Caught up in the game, you get your imaginative powers in gear and then see where they take you. Here you have to rely on your dispositions as much as in the judgmental case, though the propositions that you accept in imagining—for example, that you should remove this pie from the oven now—are not propositions that you judge.[36] The difference is that while your imaginative enterprise does require the exercise of certain virtues—virtues of inventiveness—it does not in itself require exercise of the virtue distinctively exercised in good practical judgment. Yes, you have to be a good judge relative to the propositions that you judge—for example, that you shouldn't actually eat the 'pies'—but you don't have to be a good judge with respect to the propositions that you accept in imagining. Your self-reliance in each case is a confidence that the reflexive disposition to act as if you ought to φ manifests the specifically proper virtue.

Elaborating the judgment side of this parallel, I propose that your reflexive disposition to act as if you have a preemptive reason to φ could give you a preemptive reason to φ if the disposition manifested the virtue proper to practical judgment. It is on your possession of this virtue that you distinctively rely when you judge.

IX

What, then, *is* the virtue proper to practical judgment? One temptation is to leave it at: reliability in judging what you ought to do. Of course, it's natural to maintain that such reliability has got to figure in the virtue, since to the extent that you're not reliable in judging what you ought to do on a given occasion you ought not (other things equal) to be willing to judge—at least, not without seeking advice. But I don't think it's all or even the most basic part of it. The core of the virtue is your ability to take care of yourself in appropriate ways.

When you judge—not merely *pro tanto* but 'all things considered'—that you ought to φ at t you present yourself as having met the standard for correctly concluding a deliberation in your context and as having thereby met a standard of self-care. The way to see this is to observe that if you feel disentitled to the thought that you've met this standard—say you suspect that depression or despair has got the better of you—that feeling will tend to prevent you from reaching a conclusion. Even if the feeling doesn't generate a consideration that you know how to weigh directly in the deliberation, it will typically lead you to raise the standard for closing deliberation. 'I'm not sure I can trust myself in my current state of mind,' you'll tend to think, and that thought will tend to make you especially cautious in reaching a conclusion. Of course, you may find yourself reaching a conclusion nonetheless. Then the feeling that depression or despair has got the better of you will tend to make you akratic. You'll tend not to trust your judgment, or if you do choose in accordance with your judgment, you'll tend not to trust that choice. Akrasia does not, of course, require the melodrama of depression or despair. The melodrama merely makes vivid how akrasia manifests self-mistrust.[37]

There are thus two facets to the virtue proper to practical judgment: reliability in getting it right and reliability in appropriately caring for yourself. We can say that these are two separate virtues, if we want, and that your self-reliance when you judge involves a confidence in your exercise of both. But there's a reason to resist this separation maneuver, and it derives from my basis for suggesting that reliability in self-care is more fundamental to practical judgment than reliability in getting it right.

The basis for the suggestion is that it is at best misleading to speak of 'reliability' in getting it right simpliciter, since there is no conceivable human purpose served by relying on someone simply insofar as he tends to get practical judgments right. Say you know of someone only that he makes mostly correct practical judgments. Why should this matter to you? After all, these correct judgments may be restricted to matters of no legitimate human concern. You yourself could be very 'reliable' indeed if you restricted your judgments to extremely trivial or easy issues. (Should one keep one's eyes open while typing at the computer? Should one take the day off work, hike to a remote spot, then whirl about like a dervish trying to break at least an ankle?) It is misleading to speak of 'reliability' where there is no real chance of anyone, even your own self, actually needing to rely on you. What matters is not 'reliability' in getting it right simpliciter but reliability in getting it right on relevant subject matters taken in context. But what subject matters? what contexts? My hypothesis is that you can reliably answer these questions in the sort of case at issue only if you are reliable in appropriate self-care.

Again, compare an interpersonal case. Instead of the parquet, imagine you're being led up Denali by a professional who in this tricky stage must lead your footfalls as if you were learning to walk. You trust him, but on which of his attributes does this trust lead you to rely? Obviously, you trust his competence: you take for granted that he knows where each step needs to fall. (If you get evidence that he doesn't know this, you may still need to rely on him but you'll withdraw your trust. You'll treat his interventions as giving you evidence about where to place your step, but you'll seek out additional evidence against which to weigh it.) Less obviously but crucially, you trust his attitude of concern for you. This attitude needn't express actual care for you *as you*. It need merely express a professional conscientiousness—concern for the client. The result is either way the same: he will, you trust, be appropriately attentive to your needs in respects that matter to the task at hand. (It may actually help if he doesn't care about you *as you*, since that may get him spooked by the risks you face.) The most competent guide is worthless if he isn't paying attention to your needs. You trust the competence, but more fundamentally you trust the attitude of care.

The competence and the care in this respect typically go hand in hand: too little competence makes care impossible. "But we cared," cry the incompetent parents who killed their baby by feeding it nothing but soy milk and apple juice. Well, perhaps they *felt* concern for their baby, but they didn't really care for it. Caring for it required educating themselves about its needs and seeking help from others when it failed to flourish. So with your guide on Denali: you trust him not only—perhaps not at all—to *feel* concern but to act in a way that manifests it. And so too in your self-relations. Your competence in adjudicating complex questions of how to live manifests your competence in taking care of yourself. Each competence is inseparably both an attitude and a skill.

But can't you be adequately self-concerned without being fully deliberatively competent? Though the question takes me beyond what I can argue here, I think the answer is: yes, you can. So I think you can be relevantly trustworthy while nonetheless reaching the wrong conclusion about what to do. In such a case, your trustworthiness takes the form of accurately foreseeing your practical predicament in the context of action—not merely where but *how* you'll be in that context: with what preferences and values—and deliberating in a way that is adequately responsive to that foresight. You might satisfy those conditions while nonetheless making a mistake in your deliberation—that is, while being to that extent deliberatively incompetent.[38] Then the preemptive reason created by your trustworthiness would mark a subjective rational requirement in one sense but not in another. A preemptive reason is neither a straightforwardly objective reason nor a mere requirement of self-consistency. The trustworthiness at its basis marks a subjective requirement not of mere self-consistency but of reasonable self-trust. Self-trust is not reasonable unless you are actually trustworthy. If you are trustworthy, however, you can thereby give yourself a reason to follow-through on an intention that you did not have sufficient reason to form.

It also takes me beyond the scope of the paper to observe that this virtue-theoretic approach to judgment suggests a new way to conceptualize the connection between virtue and practical reason. Instead of viewing practical reason as codifying the dispositions of practical thought that count as good character traits,[39] as if the point of practical reason were to *be* a certain way, I propose we view it as codifying the self-relations adequate to the task of judgment. The task of judgment is to constitute yourself as an authority over your actions—that is, to manifest, in the way of (PJ), the virtue of judgment. That ensures not only a practical content but a distinctively practical role for virtue, since it is only by thus constituting yourself that you can act at all.

I propose that we understand virtues as qualities codifying how we may wisely rely on people, including most crucially ourselves. In that tricky patch on Denali, you rely on your guide's patience and steadfastness; the previous day, in an unexpected storm, you relied also on his initiative and cunning. Each species of reliance is realized partly in reliance on his judgment, and each instance of the latter reliance involves reliance on both competence and care. Climbing slowly on your own, you place your footfalls with a similar reliance on your own capacities and dispositions, including those that inform your judgment: your patience, your steadfastness, your cunning, but most crucially your humility in being open to help. The role of humility in such a case—perhaps it has this role whenever we act—clarifies how one's own judgment guides one in the same way as the judgment of another. Virtue provides the key to practical reason because practical reason takes the same form whether the influence it codifies is inter- or intra-personal.

X

My principal aim in this paper has been to explain what it is for judgment to guide action by guiding choice or intention. The explanation that I offer entails a novel explanation of akrasia, but another part of its explanatory appeal is that it revealingly recodifies some core debates over the structure and function of practical reason.

As codified by (PJ), (C) and (I), judgment, choice and intention are the acts or states through which we guide our actions by making our reflexive dispositions adequate to our felt entitlements. Each act or state embodies a claim of authority, and the question is whether that authority expresses practical wisdom. It's the status of this question as non-deliberative—barring a change of mind, you're done deliberating—that creates space for the two akratic gaps. Incontinence is being disposed (a) to act as if your being disposed to act as if you have a preemptive reason *gives* you a preemptive reason, without being disposed (b) to act as if you have that reason—but also without being disposed (c) to abandon disposition (a). Weakness of will is being disposed (b) to act as if you have that preemptive reason and expecting that you'll be so disposed at the time of action, without when that moment arrives being disposed (d) to act on the reason—but also without being disposed (e) to abandon disposition (b). Each gap measures the distance between your presumption of authority and your failure to live up to the responsibilities this presumption entails. If you aren't following through on it, or aren't going to follow through, you shouldn't presume that you've settled the matter, as you do by retaining disposition (a), in incontinence, or disposition (b), in weakness. We thereby see that the exercise of practical reason is structured by norms of trust.

Having made that broad explanatory argument, I'll leave another in the form of a parting suggestion. The best way to conceptualize my view is through the analogy I've presented between interpersonal and intrapersonal trust. If I am right about agency, there is a natural explanation of how we came to be agents: imagine a caregiver instead of the dance instructor, then imagine the subject learning agency by internalizing the care. It would be theoretically economical if we could explain agency as deriving its nature from the way it is learned.

It is from this ontogenetic perspective that we can see most clearly how akrasia marks a failure of self-care. The akratic agent rejects a provision of reasons without having or seeking another to take its place. That's *irrational* but as a failure of self-care perfectly intelligible. Like rebellion in a child, akrasia is the failure to acknowledge that reasons are a limited resource and do not materialize on a whim. It is by that undeliberated acknowledgment that agency is guided by reasons, from judgment to intention to action. Each gap measures a self-relation that did not materialize, and will not rematerialize, by active willing.[40]

Notes

[1] See, for example, Sarah Stroud, "Weakness of Will and Practical Judgement," and Gary Watson, "The Work of the Will," both in Sarah Stroud and Christine Tappolet (eds), *Weakness of Will and Practical Irrationality* (Oxford: Oxford University Press, 2003). Little attention has be paid to the second akratic break, between intention and action, because until Richard Holton's recent work it had not been conceptualized as a kind of akrasia. Actually, Holton argues that it is *not* akrasia, but that's because he identifies akrasia with the other break. Holton argues that the second break is more what both philosophers and nonphilosophers have in mind when they speak of 'weakness of will,' a point I'm happy to concede. (See Holton's "Intention and Weakness of Will," *Journal of Philosophy*, 96 (1999), 241–62. As Holton notes, Amélie Rorty had earlier discussed the second break as a species of akrasia: see her "Where Does the Akratic Break Take Place?" *Australasian Journal of Philosophy* 58 (1980), 333–46.)

[2] I'll avoid talk of 'decisions,' since 'to decide' is ambiguous between the judgment side and the intention or choice side. You can decide that you should φ, a practical judgment, and you can decide to φ, a choice. The concept of decision thus obscures the fact that there are two possible akratic breaks.

[3] Why two gaps but three species of internalism? Because the first gap links two mental items (judgment and choice or intention), whereas the second links a mental item (intention) with action, which needn't be mental. Since the 'internal' link posited by internalism links a mental item with something else (which may be mental or may not), we get the possibility

of two internal links—both 'forwards' and 'backwards'—in the first gap but only one in the second. Of course, one could argue that action itself bears a species of 'backwards' necessary or non-empirical link with intention, choice, judgment or deliberation—that, for example, a bit of behavior cannot count as an action unless it is so governed. But that thesis wouldn't on its own be regarded as a species of 'internalism.' (One might wonder why not, why internalism here is a thesis about mental items. It's better to be grateful, since 'internalism' is polysemous enough.)

[4] R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," *Philosophers' Imprint*, Vol. 1, No. 3, 9–10.

[5] Richard Holton, "How is Strength of Will Possible?" in Stroud and Tappolet, *op. cit.*, 49. Again, Holton does not himself regard weakness of will as a species of akrasia. This disagreement is mostly just terminological—though I'm in substantive disagreement with Holton's account of weakness.

[6] By 'trust your judgment' (here and throughout the paper) I mean *trust yourself insofar as you've made that judgment*. I am not positing a faculty of judgment. (Parallel points apply to 'trust your choice' and 'trust your intention.')

[7] Being receptive to another's lead does not amount to relinquishing control or guidance over your actions. Your receptivity will only go so far, and *you* determine how far. If your dance lesson begins to stray toward what feels a bit too close to judo, you may withdraw your receptivity for the simple reason that that was not the sort of thing you wanted to do when you made your will receptive to your partner's lead. We might seek a theoretical rationale for this in Harry Frankfurt's conception of action as purposive behavior (see "The Problem of Action," reprinted in his *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), especially 49). Just as, in Frankfurt's example, you may remain in control of your automobile even as it coasts downhill in neutral gear and you remove your hands dare-devilishly from the steering wheel, so may you control the actions you perform when you give yourself over completely to your partner's lead. In each case your control consists in the fact that the course of events is *subject* to your intervention or adjustment, even if you never do in fact intervene or adjust anything. In the former case you control the motion of the automobile, although it is not you who even indirectly produces that motion. In the latter case you control the movings of your body, although it is not you who moves it.

[8] Expert dancers inform me that following an expert lead on the dance-floor requires a certain suppression of one's counterfactual sensitivity to evidence of untrustworthiness. But some such sensitivity is nonetheless in play: certain moves would rightly not be tolerated.

[9] 'Wondering' doesn't just mean considering the idea but regarding the idea as (that is, judging it to be) possibly correct—that is, correct as far as you know. If you *judge* that as far as you know you oughtn't to trust your judgment that you ought to φ, or your intention to φ, then you simply aren't making that judgment or forming that intention. Genuine self-mistrust is possible, and this paper is premised on that possibility. But it does not take the form of a higher-order judgment targeting your judgment or intention. (One might try to isolate a dithering species of akrasia in which you judge that you ought to φ yet 'leave open' whether you ought to redeliberate. But 'leave open' would mean either that you haven't really closed the first-order deliberation or that you have closed it but merely regard it as likely to be reopened in future. In neither case is there genuine akrasia.)

[10] If you search for reasons to rely on someone—their reliability in the past, the risks of not relying on them, etc.—you are to that extent not trusting them. We often rely on people for such reasons, and sometimes we rely on ourselves for such reasons. But that is a very different inter- or intra-personal relationship from trust. In all such cases, the reliance is fundamentally deliberated. (For more on this key aspect of trust, see my "Trust and the Second Person" (in preparation). For a full elaboration of the rather compressed argument in this paragraph, see my "Reasons and Rational Coherence" (in preparation).)

[11] Again, you can conclude and commit to deliberation without feeling certain. (Again, you need merely be as confident as the stakes in your deliberative context require.) It is a tempting

but fundamental mistake to let the possibility of lingering ambivalence obscure the roles played by such closure and commitment.

[12] It is never too late to rebel against the authority of that self, though if it is too late to stop the action the rebellion may not prove successful. We don't usually regard an agent's last-minute 'Egad, what am I doing?' as transforming the action into a mere compulsive spasm, but we do give him some credit for jumping off. (Of course, sometimes the exclamation is an exercise in hypocrisy, and we blame him both for the action and for this other action of trying to escape blame.)

[13] Holton pursues this thought at length in "Intention and Weakness of Will," *op. cit.* The debate between his position and mine depends on our different views of intention. Holton accepts the outlines of Bratman's view. In "Trust and Diachronic Agency" (*Noûs* 37 (2003), 25–51), I give grounds for preferring a view that highlights the intrapersonal side of intention that I'm emphasizing here. I'll give a briefer defense of that view of intention in section VII below.

[14] Here's an even simpler sort of case: taking dictation while tired, bored, or hungry.

[15] "The Work of the Will," *op. cit.*, 181. By 'weakness' in that passage Watson means what I'm calling incontinence; i.e. he's only discussing the first akratic break.

[16] "The Work of the Will," 183.

[17] This pretence needn't rise to the level of self-deceptive belief, so this wouldn't amount to explaining practical akrasia in terms of doxastic akrasia.

[18] The recodification will introduce ideas that are not required by my argument in sections I through V. One could accept that argument while preferring a different solution to the problems to which I'll turn in sections VI and VII.

[19] For a stab at that, see my "Regret and Responsible Agency" (in preparation).

[20] For an influential example, see T. M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998), 25–30. More recently, see Niko Kolodny, "Why Be Rational?" *Mind* 114 (July 2005), section 5; and the article by Scanlon cited in the next note.

[21] "Structural Irrationality," in G. Brennan, R. Goodin, F. Jackson, and M. Smith (eds), *Common Minds: Themes from the Philosophy of Philip Pettit* (Oxford: Oxford University Press, 2007).

[22] Depending on how we construe it in detail, such a view would deny at least one of several 'internalist' theses about the nature of reasons or practical judgment that don't have anything directly to do with my internalist thesis in this paper. In Stephen Darwall's terminology, it would deny either 'reasons/motives existence internalism' or 'reasons/motives judgment internalism,' or possibly both ("Reasons, Motives, and the Demands of Morality: An Introduction," in S. Darwall, A. Gibbard, and P. Railton (eds), *Moral Discourse and Practice* (Oxford: Oxford University Press, 1997)).

[23] I develop an account of doxastic judgment in "Judging as Inviting Self-Trust" (in preparation).

[24] See note 18 above.

[25] The result could be put to service in a compatibilist account of free will, though I'll leave that issue to one side.

[26] My usage of 'accept' resembles David Velleman's in "On the Aim of Belief" (in *The Possibility of Practical Reason*, 250, except that I've eschewed Velleman's use of the proto-judgmental term 'regard' in my gloss of acceptance. (Velleman comes very close to my dispositional formulation when he characterizes the motivational role of belief on 255.)

[27] '[Mental verb] to φ at t' is of course ambiguous, but in this paper it will never mean what could also be expressed by '[mental verb] at t to φ.' That is, the 'at t' will always express when the φing is to be done, never the time of the [mental verb]ing.

[28] Note that you needn't possess the concept of practical wisdom in order to count as acting as if you are practically wise—though you do need to possess, and be exercising, a concept codifying your sense of being reliably in the right.

[29] Note that my usage of 'preemptive reason' differs Joseph Raz's usage of the term (*Morality of Freedom* (Oxford: Oxford University Press, 1986), 42, 46). For Raz, preemptive reasons preempt other reasons that might inform deliberative judgment, whereas for me they preempt deliberation altogether. More generally, Raz posits preemptive reasons in an explanation of the relation between an authority and individual judgment, whereas I posit preemptive reasons in an explanation of the relation between a commissive attitude, informed by judgment, and follow-through on that attitude. Raz is exploring interpersonal rational relations, whereas I'm exploring intrapersonal rational relations. To the objection that there are no such reasons I reply that I am here (and in "Trust and Diachronic Agency" and "Reasons and Rational Coherence") making an argument for their existence: the argument that we cannot understand the authority of judgment without them.

[30] I'm here responding to Michael Bratman's 'bootstrapping' argument in *Intentions, Plans, and Practical Reason* ((Cambridge: Harvard University Press, 1987), 23–7, 86–7) and John Broome's reformulation of that argument in a number of more recent articles (e.g. "Are Intentions Reasons? And How Should We Cope with Incommensurable Values?" in C. W. Morris and A. Ripstein (eds), *Practical Rationality and Preference* (Cambridge: Cambridge University Press, 2001). I give a fuller version of this reply in "Trust and Diachronic Agency," section VI, and I devote nearly the whole of "Reasons and Rational Coherence" to it.

[31] Read the "at that time" in (I) to exclude such failures of self-knowledge as John Perry discusses in "The Essential Indexical," *Noûs* 13 (1979). It is, in Hector-Neri Castaneda's term, a quasi-indicator (see his "Indicators and Quasi-Indicators," *American Philosophical Quarterly* 4 (1967), esp. 93–96). I defend this account of intention at length both in "Trust and Diachronic Agency" and (from a different angle) in "How to Settle on a Shared Intention" (in preparation).

[32] Again, I defend the claim that this species of rationality can amount to acting from a reason in "Reasons and Rational Coherence," engaging in particular John Broome's polemic against the claim (see note 30 above).

[33] In "Regret and Responsible Agency" I argue that the self to whom you must appear trustworthy is not your acting self but the twice-future self that may look back in regret on the self-trust relation that you establish when you act on the intention. It cannot be a requirement on intention that you expect you'll appear trustworthy to your acting self, because you know that at the time of action your perspective may be distorted by temptation. But let's set aside that complexity here.

[34] Velleman also makes the point that imagination, supposition and pretense 'aim' at the truth in the sense that they involve imagining-, supposing-, and pretending-true ("On the Aim of Belief," *op. cit.*, 247–250).

[35] I do not in the least deny the possibility of self-deceptive or otherwise less than fully self-aware judgment or intention. It may indeed surprise you to find yourself judging that you ought to φ, or to learn that you intend to φ. You can be surprised by your own judgment or intention because you can be deceived about what you judge. The reflexivity of judging has relevance not to whether you're surprised but to what you're surprised at—that is, to the *target* of your surprise: you're surprised that you are disposed to act not only as if you ought to φ but as if you are disposed to act as if you ought to φ. You are surprised not merely to find yourself disposed but to find yourself reflexively implicated in your dispositions.

[36] Though they are, of course, propositions that you pretence-judge or make-believe-judge. For an account of imagination, or pretence-judging, inspired by this sort of example, see Kendall L. Walton, *Mimesis as Make-Believe* (Cambridge: Harvard University Press, 1990), Chapter 6. Walton's account does not, however, emphasize the aretaic dimension that I'm highlighting.

[37] Unlike Nomy Arpaly ("On Acting Rationally Against One's Best Judgment," *Ethics* 110 (2000), 488–513) I do not claim that it is possible for such mistrust of your own judgment to be rational. The issue is complicated in ways that I cannot adequately treat here. For a full treatment, see my "Reasons and Rational Coherence."

[38] If we deny that competence entails flawless performance, we can insist that the mistake does not reveal incompetence. But that will sometimes look like special pleading.

[39] For this formulation of the virtue-theoretic approach to practical reason, see Kieran Setiya, *Reasons without Rationalism* (Princeton: Princeton University Press, 2007), 69. Setiya devotes Part Two of his book to a defense of that approach as the only alternative to a discredited rationalism. I develop the thought in this paragraph in "Virtue, Internalisms, and Practical Reason" (in preparation), arguing against Setiya's dichotomy between virtue-theory and 'rationalism' by showing that the internalist and therefore 'rationalist' view that I've begun to develop here nonetheless also deserves to be classified as virtue-theoretic. The debate over reasons and 'rationalism'—ultimately a debate, as Setiya rightly notes, over several internalist theses—is therefore best pursued as a debate *within* virtue theory.

[40] Thanks to Andrea Westlund for help with the first draft of this paper, which I wrote at Clear Lake, Wasagaming, Manitoba, in June 2005. Thanks to Andrei Buckareff for raising compelling objections as commentator on a subsequent version at the Central APA in 2006. Thanks to an anonymous referee for helpful criticism of the penultimate draft in the fall of 2007. And thanks, finally, to Kyla Ebels Duggan for penetrating comments, at a conference at Northwestern in May 2008, on a shortened version of the draft that had been accepted for publication several months earlier. Despite all this assistance, I know that much is still murky and possibly muddled. As the references to work in preparation indicate, the project of which this paper is a piece remains an ongoing concern.