Forthcoming in *Philosophy and Phenomenological Research*
Dan Moller

Meta-reasoning and Practical Deliberation

Sometimes there is evidence about what we would decide to do from an *improved deliberative position*—one in which we have better information, say, or are subject to less bias, or are able to consider the relevant facts with greater vividness. I argue that in such situations we should act on that evidence, and that there are some important ethical and prudential applications for this idea. Following through with this suggestion allows us to respond to the fact that we are prone to error by making the appropriate adjustments in our decision-making. A secondary goal is to explore the neglected role of vividness in our decision-making.

Fallibilism about our practical judgments is uncontroversial: we are frequently wrong when we make decisions about what to do—often badly wrong—and everyone knows it. But efforts to characterize and respond to our lamentably imperfect decision-making have lagged. There has been some work in related fields. A few philosophers have tackled our uncertainty about moral principles and the norms that apply in virtue of that uncertainty, or argued that some moral views are better hedges against our fallibility than others.[1] In a related vein, there has been discussion of how to respond to evidence of our fallibility emanating from empirical psychology and disagreement among peers.[2] But despite this progress toward taking our fallibility as practical (and epistemic) agents seriously, much more work remains to be done, and my aim here is to contribute to this task. In particular, sometimes there is evidence

[1] See my "Abortion and Moral Risk" (draft); Ted Lockhart, *Moral Uncertainty and Its Consequences* (Oxford: Oxford University Press, 2000), and Jacob Ross, "Rejecting Ethical Deflationism," *Ethics* 116 (2006), 742-768.

[2] Adam Elga "On Overrating Oneself... and Knowing It," *Philosophical Studies* 123 (2005), 115-124, Adam Elga and Andy Egan, "I Can't Believe I'm Stupid," *Philosophical Perspectives* 19 (2005), 77-93, Tom Kelly, "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology Vol 1*, ed. J. Hawthorn and T. Szabo-Gendler (Oxford: Oxford University Press, 2005), and David Christensen, "The Epistemology of Disagreement," *Philosophical Review* (forthcoming).

about what we would decide to do from an *improved deliberative position*—one in which we have better information, say, or are subject to less bias, or are able to consider the relevant facts with greater vividness. I argue that in such situations we should act on that evidence, and that there are some important ethical and prudential applications for this idea. Following through with this suggestion allows us to respond to the fact that we are prone to error by making the appropriate adjustments in our decision-making. A secondary goal is to explore the neglected role of vividness in our decision-making.

1

Suppose that you have suddenly decided to fire a subordinate who has made an irritating but ultimately inconsequential mistake in his first week on the job. As he enters your office it occurs to you that you have not really thought the matter through very thoroughly. In fact, as he sits down opposite, you recollect a couple of similar cases several years ago when further reflection revealed that you had acted impulsively and came very much to regret what you did. In the few moments you have available, you might well reason that these past experiences suggest additional thought will lead to the conclusion that firing your employee would be a mistake, and from that you might in turn conclude that you shouldn't fire him. The first-level practical reasoning that would normally be engaged isn't available—there is no time to reexamine the reasons that count against firing—but some (rough-and-ready) reasoning about that reasoning is, and this second-level reasoning, or meta-reasoning,

is sufficient to warrant (in)action. That is, although you cannot actually reconsider whether your underling's mistake merits dismissal, you can see that if you did mull things over you would probably decide it does not, and that itself favors retention. Thus, in this kind of case we seem justified in making an inference from what we would decide after some hypothetical deliberative process to a reason we actually have. Meta-reasoning about how we would reason under more favorable conditions seems capable of telling us what to do.

But not just any such inference deserves our attention; obviously we are not interested in how our deliberations would go if we knew they would be flawed in some way, vitiated by false beliefs, irrational inferences, or the like. Notice, in particular, that it isn't generally true that we are better off simply deliberating *more*; under certain conditions a hasty guess might be better than seasoned reflection, as with Huck Finn, whose initial impulse to help Jim probably would not survive extensive and inevitably misguided deliberation. On the other hand, it isn't of much practical interest that we would be justified in acting on the upshot to ideally rational deliberative processes conducted in the light of all pertinent facts.[3] Even if we believed, as many philosophers have, that such deliberations are constitutive of what we have reason to do, we rarely or never have good evidence of what we would decide under conditions so far removed from the benighted circumstances in which we must (alas) conduct our lives. Appealing to those conditions may be useful for

[3] For a recent view along these lines, see Michael Smith's *The Moral Problem* (Oxford: Blackwell, 1994), ch. 5. See also Richard Brandt's influential discussion of "cognitive psychotherapy" in "Rationality, Egoism, and Morality," *The Journal of Philosophy* 69 (1972) 681-697. Notice that I am only claiming that such accounts are of little *practical* value. Naturally they might be useful for the kinds of *theoretical* aims that writers like Smith actually have. My comments aren't meant as criticisms of them.

theory-building, but not for agents trying to decide what to do. However, the mundane office example suggests a third alternative: when we have evidence about the outcome of deliberative processes that would be a *net improvement* on those we are in a position to conduct at present, then we have reason to act on that evidence.[4] But only when the improvement is a *net* improvement, since obviously changes in our deliberative condition might leave us better positioned in some respects and worse in others. This is the intermediate position that will be pursued here; it allows us to preserve the insight that more deliberation isn't necessarily better because of all that can go wrong in our reasoning, while eschewing the other-worldly complexities of full-blown ideal deliberation.

It would be helpful at this point if we could give an exhaustive list of candidates for improvement to our deliberations so that we could rigorously define "improved position." And, in fact, some philosophers have held that it is possible to give a brief catalogue of all that needs to go right in our practical reasoning. The usual suspects are possessing all of the relevant true beliefs but no false beliefs, not committing any deliberative "mistakes" (especially committing logical fallacies or failures of imagination), and perhaps a codicil about systematic justification or coherence as well.[5] I, on the other hand, suspect that any such list will either contain items that themselves require endless further explication ("mistakes" is very suspicious), or at the very least long and messy, especially once the role of the emotions in deliberation is acknowledged. Bias, for instance, can derail deliberation

[4] By "improvement" I mean what looks like an improvement to the agent. Since I am interested in decision-making, I am taking up the first-person perspective, not the omniscient point of view.
[5] See, e.g., Bernard Williams, "Internal and External Reasons," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), and Smith, op. cit, esp. pp. 155-161.

by getting us to respond to one but not another set of factors. Or we might be emotionally inert people who fail to care enough about certain things. Or we might care too much and find ourselves paralyzed. The possibilities for mistakes seem endless, and so too the opportunities for improvement. However, we need not settle the issue here. What's important is only that there are ways in which everyone agrees deliberation can be improved, and that evidence of the outcome of such improved deliberation can give us insight into reasons for action we might not otherwise have access to. "Improved position" will thus remain an undefined notion, though the basic idea should be reasonably clear.

We can now offer a deeper account of the intuitive claim, illustrated in the opening case, that we should let ourselves be guided by meta-reasoning. For suppose that you have evidence that an improvement in your deliberative position would yield a different outcome to your reflections. Perhaps you have evidence that if you were to remove some distorting influence or sensitize yourself to some consideration you would then arrive at a different practical conclusion. That evidence is evidence of how you would deliberate if you had better access to the reasons there are for you to act.[6] One of the goals of practical reasoning, I take it, is to try and ensure that we act on the reasons we actually have, to make it more likely what we do reflects our actual concerns and interests.[7] And that means that meta-reasoning will make us more likely to achieve the very goals that motivate practical reasoning in general, since

[6] This is not to endorse a particular metaphysical view about reasons. Even Humean internalists about reasons like Williams (op. cit.) can grant that we are fallible about the reasons we have and that we may be better or worse positioned to understand the reasons generated by our "desiderative profile." For wider discussion of deliberation from a Humean point of view, see Simon Blackburn, *Ruling Passions* (Oxford: Clarendon, 1998), especially ch. 8. For criticism see ch. 2 of Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).

[7] On the sometimes puzzling question of what the object of practical reasoning is, see David Velleman's *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), 176-178.

meta-reasoning, as conceived of here, involves discovering how we would act if were better positioned to discover the reasons we have. Sure, the evidence may be misleading; what seems to us to be an improvement might actually lead to a worse decision because of some subtle flaw that eludes us, but that's a hazard common to all practical reasoning performed by fallible agents.

This point about the nature of practical reasoning also allows us to sharpen the general thesis a bit. Not just any decision that agents face is a good candidate for meta-reasoning. There are, I believe, only a fairly limited set of circumstances in which it will be useful to reflect on our decision-making, since meta-reasoning won't always give us insight into reasons that we would otherwise be oblivious to and make it more likely that our decisions reflect our true interests. In run-of-the-mill cases, we are unlikely to benefit from the kind of self-awareness meta-reasoning implies because we aren't generally subject to large distortions in our judgment that might be removed. Also, in many cases, reflection might reveal such distortions, but we will lack evidence of what exactly an improved deliberative position would look like and how we would reason from such a vantage point (as in cases where we don't have all the facts or face up to limitations in our reasoning capacities). Or the distortions might be such that we would endorse the upshot of our deliberations even in their absence (as with some of the biases that love introduces). And, in still other kinds of cases, discussed in more detail below, barriers to good decision-making stem from weakness of the will or other cognitive malfunctions that are difficult or impossible to overcome. In the midst of a panic attack I may recognize that under more favorable circumstances I wouldn't believe myself to be dying at that very moment and would

instead go about my daily business, but that insight isn't very useful since it will probably be hard or impossible to act on. The claim is thus not that we should just generally engage in second-level reasoning, but rather that in the narrow range of circumstances in which we have evidence both that we are subject to some distorting influence in our deliberations and of how our deliberations would go under improved conditions, and when such reasoning would result in some non-trivial insight otherwise unavailable that we can *respond* to, then we should set aside our first-order judgments and act on evidence of how we would reason from an improved deliberative position.[8] Doing so, I will argue, will often lead us to change course in interesting real-world cases.

It is worth pointing out that many of the same claims apply to theoretical reasoning, though there appear to be fewer interesting applications. You might be trying to make up your mind about the truth behind global warming, and come into evidence about what you would believe if your epistemic position were improved. Perhaps you suffer from distorting biases arising from your large stake in oil companies. The opinions of other people who are relevantly like you but lack your biases arguably constitute good evidence for what you would come to believe absent those distorting influences. If people just like you, minus the stake in oil, overwhelmingly believe in man-made global warming, that gives you a reason to adopt the same view, since it is evidence of what you would think under improved

8 The issue of conflicting first- and second-order reasoning is a complex one I can only touch on here. Toward the end of the paper I pick up on just one of the thorny questions that arises—whether agents who act on second-order reasoning in a way that forces them to override first-order convictions doesn't compromise those agents in some objectionable way.

epistemic circumstances; ignoring such evidence just means rejecting an avenue of getting at the reasons for belief that are out there.

2

The cases given so far may seem slightly recherché; perhaps it isn't obvious that there are any interesting applications for second-level reasoning. But in fact, the possibilities for improving our decision-making are substantial. To see this, consider one particular aspect of deliberation that might be improved: how deeply we are engaged with the particulars of a case, which is to say how vivid our deliberations are. This matters because how *important* a consideration seems is often a function of how vividly we think about it.[9] Factors that might otherwise be brushed aside or rated relatively insignificant can assume a prominent position in our thinking as a result of being canvassed in more detail. Someone thinking about purchasing a home may know that it is located near a highway, and may have thought some about how unpleasant the noise will be. But despite having all of the relevant information, spending some time actually experiencing the noise, or simply thinking about it in greater detail, may affect the buyer's decision by raising the salience of that factor. The same is true of deciding to have children. It is one thing to read a summary paragraph of the experience of raising a family, another to spend a week witnessing a couple of toddlers in action. That week may convey no new information—perhaps you already knew that toddlers require round-the-clock care and are given to

[9] What I call "vividness" is related to what is sometimes called "moral imagination," and the benefits of experience or deeper engagement are similar to those often attributed to reading novels. But "moral imagination" usually refers to seeing the existence of certain considerations one might otherwise be blind to, not their level of importance. See, e.g., the introduction to Martha Nussbaum, *Love's Knowledge* (Oxford: Oxford University Press, 1992).

caterwauling. But it may nonetheless cause you to reconsider by getting you to see how important those considerations are. As cases like these bring out, the quality of our decision-making depends not only on possessing relevant information and avoiding inference errors or other formal mistakes, but on correctly assessing the relative significance of various considerations, which in turn depends on a variety of psychological operations—thinking hard and long, teasing out details, etc.—that "vividness" is meant to capture.[10] (More on vividness and some objections to this account later on.)

Now take famine relief and deliberations about whether to render aid and if so to what degree. Most of us see some reason to help those going hungry, but typically ascribe fairly little weight to that reason in comparison with promoting our own or our families' interests. But there is powerful evidence to suggest that our deliberations would be swayed if we were more deeply and vividly engaged in the relevant particulars. Suppose the money we are considering donating is for a village well. What would be involved in maximally vivid, fully imagined reflection on the relevant particulars? Ideally, we would want to acquaint ourselves with the villagers' situation, get to know their present condition, their prospects without a well, familiarize ourselves with the kinds of goods they do or might enjoy under flourishing conditions, understand what death by dehydration looks and feels like, and so on. Arguably, we might even want to get to know the villagers as individuals—their personalities, hopes, fears, life-histories, and interrelations. This is

[10] There are many complexities that I cannot explore here. One is that importance might not just be a matter of weighting a factor appropriately. A deontologist, for instance, is unlikely to think that his discovery of the importance of a certain right is the discovery that some consideration hasn't been given enough weight.

the kind of detail that will allow us to fully appreciate what the loss of the villagers would mean; it would put us at the opposite end of the spectrum of someone making decisions at a desk on the basis of a summary description. Absent these kinds of details, we are dramatically less likely to grasp the importance of our decision and reflect on all of its consequences. Just as we are unlikely fully to appreciate the importance of highway noise or caterwauling in the earlier examples, so we are unlikely fully to grasp the loss of the villagers without delving into these kinds of particulars. Of course, gathering and fully engaging with all of this information would require a prohibitive investment; perhaps it could only be accomplished by reviewing weeks worth of in-depth video footage documenting these various circumstances, supplemented by lengthy on-site visits. If the case were a more involved one—perhaps a decision that would effect thousands of lives—the obstacles to full deliberation would increase exponentially. But the claim at this stage in the argument isn't that we ought or can do any of this, it's merely to characterize what fully vivid reflection would look like, and therefore what (maximal) *improvements* in our deliberations would look like. We may not be able to reach the vertiginous heights of vividness represented here as lofty ideals, but we can acknowledge them *as* ideals, and recognize that movements in that direction would constitute improvements.

The next point is that the evidence suggests that such improvements would result in our concluding that there was a far stronger reason to render aid than we usually suppose. We are more likely to decide we should render aid when we have the condition of those in need emblazoned on our minds than when we have

considered them only superficially. This is probably not very controversial, but, in addition to the noisy house- and children-type examples, we can also look to advertising as a natural experiment confirming this claim. Advertising not only gives people new information but attempts to raise the salience of products already familiar. It works, in part, because in forcing us to dwell on considerations favoring the purchase of a product, advertising makes it more likely that we decide to buy. Especially relevant, of course, are advertisements ("appeals") from relief agencies. These are effective in producing donations even for perfectly familiar needs simply by emphasizing through pictures and narratives details we otherwise wouldn't concern ourselves with. We can infer that the improvement brought about by additional vividness in the famine-relief case would lead us to discover a stronger reason for aid. And since the additional vividness would mean an overall improvement in our deliberative position, we should take that subjunctive truth to constitute a powerful reason for acting, albeit one we are unlikely to have thought much about. To put the same point in reverse, to the extent we find ourselves thinking our reason for giving to famine relief is weak or only moderately strong, we have evidence that this is the upshot of a deliberative process that is defective or at least suboptimal in important ways, and that if these defects were remedied the result would be discovering a stronger reason to give.

Advertising points to a wrinkle, however, that is special to meta-reasoning about vividness and should be ironed out here. Notice that there are clear examples in which engaging with particulars is inimical to sound judgment. Imagine, if imagination is required, that you are on a diet, yet considering eating ice cream.

Persuaded that enhanced focus on particulars improves deliberation you begin to contemplate slowly sliding your tongue along a cool, sweet scoop of coffee ice cream. Parallel to getting to know the dying villagers, you imagine each individual lick of ice cream in protracted detail. But as fat people everywhere know, this way of deliberating does not lead to sound judgment. The same goes for generals agonizing over the gory details of each soldier's death (reviewing autopsy reports, photos of battle-wounds, etc.) which might make it impossible to order troops into battle even in a just cause. Part of the insidious function of advertising and propaganda is in fact precisely to warp our judgments by getting us hyper-attentive to select particulars (e.g., slow-motion images of ice-covered beer or incessant references to past national tragedies).

These examples bring out the importance of symmetry in our reflections. Dwelling on the details of certain features of a case will sensitize us to the importance of those features, and naturally if we sensitize ourselves to some features but not others we will misjudge the relative significance of the various considerations that are at issue. To avoid this mistake, we must sensitize ourselves to all of the considerations in roughly equal proportions. If the dieter thinks hard about remaining fat—examines pictures of his profile and reads relevant medical studies—and the general reflects carefully about what will happen to the civilians he is defending from unjust aggression, then there is no reason to suppose the result will be a failure in deliberation. Of course, the same holds true in the famine-relief case: we should envision the improved deliberations being symmetrical by imagining them to involve thinking hard about all that we might do with the money that would otherwise go

toward the well. (Perhaps the on-site visit to the dying villagers could be conducted from the SUV one would otherwise purchase.) But whereas this symmetrical corrective seems likely to dampen the dieter's craving for ice cream, at least to the extent it ought to be dampened, I conjecture that the same is unlikely to be true of famine relief. In the status quo, we think little about the details of the effects of not helping the villagers and a great deal about the SUV we wish to buy. Because we already tend to think hard about the non-aiding option, adding the same hard thought to the other side seems likely to result in discovering a stronger reason to give. Perhaps there are absolute limits to this reasoning about vividness. Perhaps wallowing in enough gory details would produce some bizarre reaction or prove absolutely paralyzing.[11] But in that we case we are no longer considering *improved* deliberations; genuine improvements will emerge, presumably, from heightened and symmetrical sensitivity to the relevant features of the case, but probably not outlandish absolute degrees of sensitization.

But this point about symmetry may not get at the worry. The real concern, it might be said, is that certain features of the moral landscape are more likely to increase their influence on us through vivid deliberation than others. "Hot" considerations like the pains or pleasures certain people experience seem more and more important as we focus on them more intensely and graphically, while "cold" considerations like abstract principles of justice or the Categorical Imperative do not. When I contemplate illegally transferring a money from Rich's account in order to aid Poor, focusing on the intense, sufferings of Poor may make that choice harder and

[11] See the criticisms of full-information accounts of rationality in Thomas Carson and Paul Moser, "Relativism and Normative Non-realism," in Paul Moser (ed.) *Moral Relativism* (New York: Oxford University Press, 2001), 287-304, pp. 293-295.

harder to resist, while the abstract principles of justice that count against doing so may not symmetrically increase their pull on me. (Assume that Rich is entitled to his wealth, and that he won't feel a moderate loss—focusing on him just conjures up images of a champagne-drinking playboy.) Vividly entering into particulars thus may seem to introduce a bias in favor of certain substantive considerations rather than others, and so in cases like famine relief the deck may seem to be stacked against the cool, abstract considerations of justice that might, for all I have said, show that aid is not required.[12]

In response, consider the possibility of "leveling down" in order to restore the deliberative balance between hot and cold considerations.[13] Suppose that we have already come to learn about the suffering of Poor in great detail, and that we are now in danger of neglecting the importance of the abstract considerations of justice. One way to improve our deliberative position would be to start ignoring the plight of Poor—avoiding conversations about him, hiding pictures, etc.—and waiting until he's nearly forgotten. At that point, the abstract considerations of justice will be able to assume a greater relative prominence in our decision-making. But this seems wrong: ignoring morally important issues like the plight of Poor is a mistake; Poor could justly accuse us of failing to take in the full depth of his suffering, which is surely something that should at least feature in our decision-making, even if it is ultimately outweighed. What we ought to do in this situation is rather to "ratchet up" by focusing more attention on abstract considerations of justice. The above-objection seems to presuppose that this is difficult or impossible, but it's hard to see why.

---

[12] I am grateful to an anonymous reviewer for pressing this point.

[13] Compare the leveling-down objection to egalitarianism. Derek Parfit, "Equality and Priority," *Ratio* 10 (1997), 202-221.

Taking Kant's cerebral, Teutonic approach to morality as the limit case, notice how easy it is to improve our sensitivity to the importance of the Categorical Imperative, particularly in its second formulation. In the case of Rich and Poor, the more we focus on the way in which stealing from Rich is to *use* Rich, to treat him merely as a means to attaining some other aim, how utterly incompatible this is with his dignity as a person, the more intense will be our appreciation of the significance of the Categorical Imperative (such as it may be).[14] So ratcheting up seems both possible and desirable compared to leveling down in the case where we have already attained vividness. But the same point holds prospectively: increased vividness can improve our deliberations because in the status quo we will by construction be failing to take genuine moral considerations as seriously as we might, and because we can ratchet up our sensitivity to abstract considerations of justice at the same time so as to avoid undue imbalances.

But this line of thought may trigger a more general objection to what I have been claiming. The famine-relief case leans heavily on engaging vividly with the particulars, and we might wonder whether these are morally relevant at all. Why should it matter that we know what the names of the individual villagers are? All that such knowledge may seem to accomplish is producing an emotional response. And these emotional reactions to fine-grain minutiae, while useful for *motivating* agents, aren't themselves capable of telling us what we ought to do. That is determined by abstract principles, the objection continues, whose application is independent of the

[14] Kant, of course, would be horrified by the thought of a moral decision being motivated by anything but duty. But Kant was clearly able to produce a deep, vivid awareness of the importance of duty, as revealed, for instance, in the famous Ode to Duty ("Duty!—Great and mighty name that embraces nothing that is loved or ingratiating, but demands submission," etc. etc.), *Critique of Practical Reason*, 1.1.3.

details I have been emphasizing. Kantians and consequentialists who think that moral deliberation primarily involves applying principles to cases might be especially attracted to an objection of this sort. In fact, they might go on the offensive and offer a positive argument. It begins with the point made earlier, that not all subjunctive truths linking changes in our deliberative environment with changes in our judgments matter; only changes that represent genuine improvements are significant. But this means that we can only evaluate the significance of evidence about how our judgments might change in light of some antecedent moral standard; moral principles must be logically *prior* to any such evaluation. Otherwise we will have no way of determining whether the fact that we would judge there to be a strong reason to do X under deliberative conditions Y is genuine evidence of the reasons that are out there or is simply misleading. And if we are already in possession of moral principles that tell us what matters morally, then there is no need for minute scrutiny of particulars that elicit affective responses; we can simply apply the principles directly to determine what we must do. Whatever benefits vivid contemplation of the particulars might yield are strictly motivational, the argument concludes.

The objection is right that getting to know the villagers is unlikely to reveal new information that any plausible moral theory would tell us was relevant to moral principles. The mistake lies in assuming that all there is to deliberation is robotically applying abstract principles to high-level descriptions ("strangers will die unless I send them money").[15] As I have been stressing, deliberation also involves trying to assess how important various considerations are, and here vividness is of tremendous

[15] As leading Kantians have themselves stressed. See chapters 4 and 7 in Barbara Herman, The Practice of Moral Judgment (Cambridge: Harvard University Press, 1993), and Onora O'Neill, *Toward Justice and Virtue* (Cambridge University Press, 1996), esp. ch. 1.

value. There is simply no way for us to fully weigh the significance of large numbers of people dying without confronting these details. This judgment about importance may well involve or on some theories even be constituted by an emotion, but that is something we can remain neutral about. What matters is only that engaging the details improves our deliberative position by facilitating judgments of importance, however that psychological process is interpreted. This leaves us with the positive argument—that we cannot be sure that changes in our views under new deliberative conditions would represent improvements without antecedent moral knowledge. The argument succeeds in showing that we must have *some* moral knowledge, but not that such knowledge must be of the sort that renders evidence about deliberation in improved circumstances redundant. What is needed is only knowledge that certain deliberative conditions are better than others; there is no need to have an algorithm that renders deliberation moot. Even someone completely at sea when it comes to moral theory might know that deliberation will go better if it is informed by the facts or conducted sober rather than drunk. He can know this simply by knowing very general facts about what kinds of things are conducive to good decision-making.

Of course, some moral theories might leave less to deliberation than others; there is a spectrum ranging from perfect algorithms to theories that offer only a few rough guidelines. And in a give case (like famine relief) a particular theory might not leave agents enough room for meta-reasoning to be important. But we should bear in mind, first, that the point was never supposed to be that meta-reasoning would help any agent using any moral theory in any situation, and, second, that there is more to practical reasoning than morality. In fact, the example of famine relief is especially

useful for making this point. Many people don't give because they feel that they are morally required to do so, but because they simply feel sorry for those they aid or because they are outraged that anyone must die "like that." These might be very broadly classified as moral reasons, but they certainly needn't emanate from a moral theory. And yet they still might be reasons the agent cares about deeply and whose status might emerge more clearly under improved deliberative conditions.

3

Later on I will take up additional objections, but for now let us turn to prudential reasoning and a different kind of case. In the initial office example, meta-reasoning is useful because there is a time-constraint that prevents you from thinking things through more thoroughly, while in the famine-relief case the problem is that attaining maximal vividness would require a prohibitive investment. But sometimes meta-reasoning is useful simply because we find certain errors in our first-level deliberations extremely difficult to avoid, even when there are no external barriers to making the relevant improvements. Consider the important finding from empirical psychology that people consistently fail in their predictions of the impact future events will have on their subjectively interpreted happiness (they fail in their "affective forecasting").[16] Through a variety of psychological mechanisms, including a tendency to ignore hedonic adaptation to adverse events and to focus solely on and exaggerate the significance of the adverse event itself, we underestimate our own

[16] *S*ee, e.g., Daniel Gilbert and Jane Ebert, "Decisions and Revisions: The Affective forecasting of Changeable Outcomes," *Journal of Personality and Social Psychology* 82 (2002), 503-514, and Daniel Gilbert et al., "Immune Neglect: A Source of Durability Bias in Affective Forecasting," *The Journal of Personality and Social Psychology* 75 (1998), 617–638. For an accessible overview of the literature, see Daniel Gilbert, *Stumbling on Happiness* (New York: Knopf, 2006).

resilience in the face of even great crises. We are consequently often wrong, sometimes dramatically so, about the effect that events like the loss of a spouse, or sustaining a spinal-cord injury, or suffering a major career setback will have on us; these events are typically far less damaging to us than we tend to predict.[17] This is now a fairly robust finding replicated by many researchers. To focus on a specific case, it turns out that whether or not assistant professors receive or are denied tenure has little impact on their happiness. In a study of over a hundred academics, multiple psychometric questionnaires designed to determine participants' levels of happiness failed to find any significant difference between those who received and those who were denied tenure, either in the first five years after tenure (aggregated), or the next five. If the more detailed studies of bereavement are any indication, being denied tenure causes a minor blip lasting a few months, and thereafter has little effect on most people's subjective well-being.

Let us assume that this finding on tenure is accurate, as is suggested by the numerous similar findings in other domains mentioned earlier. How is this relevant to decision-making? Suppose that you must determine how much importance to place on receiving tenure—perhaps in the course of making trade-offs between family life and academic work.[18] Despite the possibility that some objective value, independent of psychologically interpreted happiness, is assigned to being tenured, it is likely that much of the rationale for pursuing a career-goal like tenure depends on

[17] See my [reference omitted] for a more thorough investigation of the significance of our resilience to traumatic events, focused on bereavement. For experimental evidence, see, e.g., George Bonanno et al. "Resilience to Loss and Chronic Grief: A Prospective Study from Preloss to 18 Months," *The Journal of Personality and Social Psychology* 83 (2002), 1150-1164. On spinal-cord injuries, see Philip Brickman, Dan Coates, and Ronald Janoff-Bulman, "Lottery Winners and Accident Victims: Is Happiness Relative?" *The Journal of Personality and Social Psychology* 36 (1978), 917-927.

[18] Notice that a successful family life is associated with higher levels of happiness. See Richard Layard's *Happiness* (New York: Penguin 2005), 63-66.

its connection to making us happy. This means that making undue sacrifices to career is simply a mistake. Failing to achieve a major career goal will have very little impact on us, and consequently is not worth pursuing at the cost of other more important goods, like successful relationships. The problem, however, is that we find it immensely difficult to adapt our decision-making to take into account the information contained in the psychological studies. Many people who know of the research continue to make decisions on the assumption that the loss of a loved one would leave a lasting mark, and that being denied tenure would have a significant negative effect on their lives. The reason for this is obvious: many of the psychological mechanisms responsible for our mistaken affective forecasting are ones we find it extremely hard to overcome. The success of some of them even depends on our being oblivious.[19] Take rationalization. One reason being denied tenure isn't so bad is that we tend to rationalize that its denial isn't a reflection of our mediocrity, but rather a reflection of others' failure to appreciate our genius, of sheer misfortune, of a cabal against us, and so on. This process protects us from counterproductive despair, but it only functions if we are unaware of it—we cannot successfully rationalize that our enemies conspired against us while simultaneously acknowledging that we are merely rationalizing. Other processes, such as our tendency to focus on the painful event and neglect all of the other events that will soon succeed it, are simply difficult to correct for. For these reasons, we tend to fail to incorporate information about things like tenure's effect on us in our decision-making, and this is where second-level reasoning can help. It tells us that if we are tempted to think that tenure is important enough to pursue at the cost of family

[19] Gilbert et al., "Immune Neglect," 634-635.

relationships, we are probably making a mistake. From a better vantage-point, one free from various cognitive distortions, we would decide to pursue a more reasonable work-family balance. We should reconsider decisions about the importance of our work in light of evidence that those decisions have been compromised by the inadequacy of our deliberative position. Meta-reasoning in this type of situation can thus help us to overcome our inability to incorporate certain kinds of information into our decision-making.

This claim can seem trivial if it is misunderstood. Recall the dieter, and suppose that he has decided to ignore his first-level reason to avoid eating a giant ice cream sundae. What he needs, it might be said, is to act on the reasons he has to lose weight; telling him that from a better deliberative position he would choose not to eat the sundae, while true, isn't helpful at all. What this bring out is that not just any failure to act on our reasons is a good candidate for second-level reasoning. The dieter's problem is not that there are reasons for acting that he has trouble recognizing (as the famine-relief giver has trouble gauging the true importance of the well); his problem is probably just weakness of will. It's unlikely that he labors under some delusion that eating ice cream really is the right thing for him to do (it's obvious that dieters shouldn't eat sundaes) but for all that it's sometimes just very hard to resist what we know to be temptations. By contrast, someone who has decided to sacrifice family relations to work is much more likely to be failing to see the reasons they have to act. For the reasons given above, even if they know of the psychological studies, they are apt to conclude that achieving tenure is deeply important to their well-being. Second-level reasoning is useful in this context because by forcing us to think about

what we would decide to do if we were better positioned, it can help us to achieve an insight that might otherwise be unavailable.

Of course, even with that insight into the reasons we have for ignoring appearances, we might still make the wrong decision. But at least meta-reasoning does give us the insight, which gives us a fighting chance. And there is in fact empirical evidence to suggest that second-level insight makes it more likely that we will improve our decision-making. Researchers have found that our tendencies toward exaggerated discounting of future costs and toward procrastinating—both phenomena we tend to succumb to at the first level—are better handled once subjects have a self-awareness of these tendencies and make conscious, second-level attempts to combat them.[20] Though these studies don't involve quite the same degree of delusion as the tenure case, they nonetheless suggest at least in a tentative way that self-awareness and the attendant second-order reasoning can facilitate better decisions.

And, once again, the same style of argument applied to the tenure case can be marshaled to many others as well, for instance medical decisions that require us to balance what seem to us major calamities against the risk of premature death. A diabetic who has contracted gangrene may be tempted to delay amputation, but the evidence suggests that losing a leg would have comparatively little effect on the person's happiness. Even if he finds this impossible to see or accept, a little meta-reasoning shows that amputation may nevertheless be the right decision.

---

[20] For references and discussion of these studies, along with a fascinating application to sex, see Dan Ariely and George Loewenstein, "The Heat of the Moment: The Effect of Sexual Arousal on Sexual Decision Making," *Journal of Behavioral Decision Making* 19 (2006): 87-98.

But the tenure-type case also confronts us with two complications. First, the fact that most people commit a certain mistake doesn't show that a particular agent is, and, second, even if an agent is making the relevant mistake, that doesn't show that avoiding the mistake would affect his judgment. Suppose you are making the career-family decision. Are you exaggerating the horror of being denied tenure? The evidence suggests that you probably are, but that is no guarantee. And even if you are, perhaps your fundamental values are such that no amount of correcting would yield a different verdict. Some people might just be incapable of carrying on without being a tenured academic, and in that case second-level reasoning is silent. What we are likely to end up with is something like, "There is strong but not absolutely dispositive evidence that A's deliberations could be improved, and there is strong but not absolutely dispositive evidence that in those improved conditions A would do X rather than Y." Nonetheless, we should at the very least employ meta-reasoning to guide us in how seriously we take the possibility of error. Even in the face of a very strong sense that tenure trumps family, if the second-level considerations are strong, we should exercise extreme caution by consulting with others and making conscious efforts to overcome the various biases or illusions we know we are susceptible to. But more than this, we should, when possible, hedge: if the action is not an all-or-nothing affair (as in the famine-relief case, where money is continuously variable) we should move in the direction the meta-evidence points to. Finally, in an all-or-nothing case, the agent should try and estimate the conjunctive probability that he has committed some deliberative mistake and that his decision to act or not act depends on that mistaken judgment. If this is, on balance, likely, obviously the agent should

reconsider his choice. I don't pretend making that sort of evaluation would be at all easy, but in individual cases the answer may be clear. Certainly in the case of famine-relief, if we grant that vividness of the sort described makes for an improvement, then there doesn't seem much doubt that most agents' choices do rest on deliberative mistakes.

4

The claims I have been developing might be summarized as follows:

1. Sometimes we have evidence of how improvements in our deliberative position would affect our deliberations.
2. When we do, that evidence provides us with a reason for action—a reason we might not otherwise have access to.
3. There are important real-world applications for this idea, since there is evidence about how various standard moral problems would be affected by improvements in our deliberative position.

On the view I have been sketching, in deciding what to do we should not only cast about for reasons that are directly available to us, we should consider what reasons we might be (partly) blinded to because of shortcoming in our deliberative processes, and what evidence there might be about how our deliberations would be affected by removing those impediments. Let me conclude by rehearsing at length something that might worry us about this whole approach.

Acting on second-level evidence may mean acting for reasons that we cannot directly access, and that can seem deeply disturbing. In "easy" cases this may not create a problem. If you see a square tower in the distance, there is nothing unsettling about deferring to someone who is much closer and therefore better positioned, even if his judgment that it is in fact a round tower conflicts with yours. But in this kind of "thin" case there isn't much room for deliberation as such—there is no issue of weighing and reasoning about the relevant features of the case. What are worrisome are the cases where we don't just act on evidence we cannot directly confront, but where we act on reasoning we cannot ourselves reproduce. Take the case of a general deciding whether to send troops into battle. If he has lost friends in battle recently and that risks distorting his judgment, then the claim is that he ought to consider and possibly act on evidence of how he would deliberate under improved conditions—evidence that may very well be available in the form of the judgments of others who are similarly positioned but for having just seen their friends die. To sharpen the case, suppose that the general's judgment differs from that of the fellow-generals he consults. Here the general is being asked to make life-or-death decisions on the basis of reasons whose validity he cannot himself assess, perhaps not even understand, and which strike him as flat wrong. The same is true in the famine-relief case, albeit less dramatically. Saying why exactly this is objectionable is not easy, but one way of making the point is to say that the agent is being asked to fundamentally compromise himself by acting contrary to his own best lights. His actions aren't ones that he can endorse in virtue of his own reasoning, and that represents an unacceptable abdication of his responsibilities as an agent.

There is no question that there is something distressing about the prospect of making major decisions without being able to understand or endorse all of the judgments that serve as inputs to one's deliberations.[21] The problem is that there is no acceptable alternative. Failing to act on second-level reasoning means, by definition, acting on the basis of flawed deliberations when there is evidence of how improving those deliberations would affect the outcome. To return to the general, he presumably cares about discharging his duties as a soldier with integrity, so why shouldn't he pursue every available opportunity to improve his decision-making? It is true that in that pursuit he may cut himself off from ultimate comprehension of the reasoning that leads to the conclusion to send in the troops, but accepting that limitation would in this circumstance express precisely a commitment to his deepest values and identity as a soldier, not any sort of compromise or abdication. He can see clearly the evidence of how his improved deliberations would go, and he can conscientiously endorse that higher-order consideration as a reason for acting, even while it remains obscure to him, perhaps, how the first-order considerations generate the conclusion they do. There's no denying that the switch from acting on first- to second-level considerations might be difficult, but not doing so would often be culpably negligent and doing so can be defended by emphasizing that the shift is one of levels, not of demanding that agents stop reflecting on their own decisions.

[21] Notice that the same problem arises in many of the other context mentioned at the outset, e.g., in acting/omitting on the basis of considerations of risk or disagreement, or in deferring to experts.