This is a pre-print. The final version can be found in *Philosophy and Public Affairs* 42 no. 1 (Winter 2104): 3–41.

Reflection and Responsibility

Pamela Hieronymi
hieronymi@ucla.edu
November 19, 2013

A common line of thought claims that we are responsible for ourselves and our actions, while less sophisticated creatures are not, because we are, and they are not, self-aware. Our self-awareness is thought to provide us with a kind of control over ourselves that they lack: we can reflect upon ourselves, upon our thoughts and actions, and so ensure that they are as we would have them to be. Thus, our capacity for reflection provides us with the control over ourselves that grounds our responsibility.

I will argue that this thought is subtly, but badly, confused. It uses, as its model for the control that grounds our responsibility, the kind of control we exercise over ordinary objects and over our own voluntary actions: we represent to ourselves what to do or how to change things, and then we bring about that which we represent. But, I argue, we cannot use this model to explain our responsibility for ourselves and our actions: if there is a question about why or how we are responsible for ourselves and our actions, it cannot be answered by appeal to a sophisticated, self-directed action. There must be some more fundamental account of how or why we are responsible.

I will replace the usual account with a novel but natural view: responsible mental activity can be modeled, not as an ordinary action, but as the settling of a question. This shift will require abandoning the tempting but troublesome thought that responsible activity involves discretion and awareness—which, I argue, we must abandon in any case.

1. Overview: The Common Line of Thought and a Response

I begin by roughly sketching the common line of thought together with my response.

We are, it seems, responsible for our intentional actions, if we are responsible for anything. Intentional action provides a kind of paradigm case of responsible activity. Intentional action also seems to involve, at least in *its* paradigm instances, a certain sort of "having in mind." In the

paradigm cases, we act intentionally by first deciding what to do and then doing what we decided. We act, it seems, by being the cause of our own representations.[1]

This "having in mind" involved in decision or intention provides, I believe, much of our sense of our control over our own actions. We *control* our actions, it seems, because, or insofar as, we can think about what to do and then do what we take to be worth doing. Our sense of control over our own actions thus involves both a certain kind of *awareness*—we have in mind what we intend to do—and a certain kind of voluntariness or *discretion*—we can decide to do whatever we think worth doing. It is very natural to think that this sort of control, the kind that, in its paradigm instances, involves both discretion and awareness, is not only a ground for but also a condition on our responsibility for our intentional actions: that we are responsible because we enjoy such control, and that, if we lack it, we cannot rightly be held responsible.

However, if we start with the thought that, whenever we control a thing, we do so by reflecting upon that thing, deciding how it should be, and then bringing about that it is that way, we run into difficulties when we reflect upon our lives. It seems you should be able to reflect upon your life, decide how it should be, and then, with some effort and luck, bring it about that it is that way. However, when we reflect upon our lives, we might notice that each decision we make, and each thing we do, can be adequately explained by conditions in place prior to it. And so it might seem that we do not control our lives, after all: the future, it seems, will be explained by the past, and, since there is nothing we can do, now, to change the past, it seems there is nothing we can do, now, to change the future.[2] And so, if we start with the thought that we control a thing by reflecting upon it, deciding how it should be, and then bringing it about that it is that way, reflection on the

---

[1] In particular, by being the *intentional* cause of our own representations. The simpler formula is Kant on desire: "The *faculty of desire* is a being's *faculty to be by means of its representations the cause of the reality of the objects of these representations*." Immanuel Kant, *Critique of Practical Reason*, trans. and ed. by Mary Gregor, Cambridge Texts in the History of Philosophy (Cambridge: Cambridge University Press, 1997). 5:9n.

[2] This is one way threat to freedom appears. There are others. I treat the topic in more detail in Pamela Hieronymi, "The Intuitive Problem of Free Will and Moral Responsibility," (in progress).

course of history will erode our sense of control over even our own intentional actions. A sort of threat appears, sparking the free will debate.[3]

Parties to that debate can be aligned, very roughly, on an axis. At one extreme are those, like Roderick Chisholm and, before him, Immanuel Kant, who believe that our autonomous activity is not fully explicable by facts outside of us;[4] we are the ultimate source of our actions, which are not determined by any of our contingent psychological features.

At the other extreme lie those who think that responsibility is ultimately for *being*, rather than for *doing*. We are responsible for our actions because they are explained by and so reveal our character, or our contingent psychology, but we need not exercise any ultimate control over that character to be responsible for it. We are responsible for it simply because we *are* it. R. E. Hobart long ago provided a particularly eloquent defense of this position, grounded in an account of what it is to be responsible. To be responsible for an action, he explained, is simply to be open to certain sorts of character assessments on account of that action. Thus, to be responsible for an action, that action must accurately reflect one's character.[5] Often enough it does, and so often enough we are responsible. The fact that we are not ultimately self-created, or that our actions have sources outside of us, poses no difficulty. On the Hobartian view, to be responsible is simply to be, and to act as, yourself.[6]

---

[3] It is typically identified as the threat of causal determinism, or sometimes as the "causal thesis," but I believe that the intuitive threat to our sense of freedom can be generated simply by appeal to the fact that our choices and actions are adequately explained by facts that pre-date any of our thoughts. I treat this in more detail in Ibid.

[4] (other than, perhaps, by the demands of Reason, which, it is argued, are not constraints on freedom, but rather a condition for it, and not a fact alien to us, but given by the nature of our will)

[5] See R. E. Hobart, "Free will as involving determination and inconceivable without it," *Mind* 43(1934). On this view, character is the object of assessments, and actions reveal character insofar as they are chosen—because one's choices, it is presumed, reflect one's character. But there is no further, similar requirement, that one have chosen one's choices. At that point, the requirement looses its *raison d'etre*.

[6] Hobart represents one extreme of what Susan Wolf calls "Real Self" views. I have just, in effect, re-traced her distinction between the "Autonomy View" and the "Real Self View." Susan Wolf, *Freedom within Reason* (New York: Oxford University Press, 1990). She argues that many positions attempting to find middle ground are what she calls Real Self views, falling prey to the objections that can be raised against Hobart.

Each extreme seems unsatisfying. The first seems to require positing some or another in-principle mystery—either a constraint on our explanations where it seems that none exists or else something like a noumenal self or a soul, whose decisions, though efficacious, are (awkwardly) not (wholly) explicable in terms of the contingent psychology of the empirically given subject.[7] The second avoids the mystery by giving up the claim our responsibility is grounded in and conditioned by some form, or at least possibility, of activity or control. But that seems too steep a cost.

So there are a variety of middle positions, which try to show how we are in some sense in control of the selves for which we are responsible. The most influential of these middle positions, over the last four decades, belongs to Harry Frankfurt, and the dominate feature of most views attempting to avoid the extremes of Hobartian appeal to character and Chisholm's immanent causation is an appeal to reflection or hierarchy.[8]

It is not hard to see why this might be. By appealing to reflection, or hierarchy, we seem to recreate the sense of control—the awareness and the discretion—of intentional action. The one who reflects is aware of and exercises discretion with respect to that upon which she reflects. Thus it seems, if we can reflect upon and change *ourselves*, we enjoy a kind of control over ourselves similar to the control exercised in intentional action. Less sophisticated creatures cannot gain this

---

[7] And either the choice is, ultimately, inexplicable, or else it be explained by (something like) Reason. If we take the latter path, it is hard to see how there could be such a thing as responsible unreasonableness. If we take the former, we generate two problems. First, as Harry Frankfurt pointed out, it now seems impossible to know when a choice has been made. Harry Frankfurt, "Freedom of the will and the concept of a person," in *The Importance of What We Care about* (Cambridge: Cambridge University Press, 1988), 23. Second, it leaves us holding an actual, empirical subject responsible for a choice that seems not attributable to her.

[8] In his very early paper, Harry Frankfurt provided an example that was meant to show that alternate possibilities are not required for moral responsibility: it is not true, he said, that one is responsible for an action only if one could have acted otherwise. Rather, whether one is responsible for an action turns on whether your action is to be explained by your choices, or, perhaps, by appeal to what you really wanted (in some yet-to-be-determined sense of "really"). (See Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66(1969).) In Harry Frankfurt, "Freedom of the will and the concept of a person," *Journal of Philosophy* 68, no. 1 (1971)., he begins what later becomes an extended attempt to say what it is to "really" want to do something—what is required for a choice, or for one's will, to be one's own. In this second article, the most salient feature of the developing view is the appeal to hierarchy, or self-reflection.

kind of reflective distance, and therefore they are not responsible for their thoughts or their actions in the way we are.[9]

I believe this reflective strategy is mistaken. My basic reason for thinking so is rather simple. The strategy appeals to reflection as a way of securing control over ourselves. But *merely* being able to reflect upon a thing does not provide one with control over that thing. (Think of Kant's creature from Part I of the *Groundwork*, endowed with only theoretical reason, able only to contemplate its happy state while instinct controls its movements.) If one is to control something of which one is aware, one must also be able to *change* that thing—in particular, to bring it to accord with one's thoughts about how it should be. However, insofar as the reflective strategy secures our control over ourselves by appealing to the fact that we can reflect upon and change ourselves, it has, it seems, secured our control over ourselves by appeal to a self-directed action. But this will not do. If there was a question or problem about how or why we are responsible for our intentional actions, we cannot answer it by appeal to a self-directed intentional action.

2. First Reply of the Champion of Reflection

The champion of reflection will object that her position is here caricatured. I will consider two replies. First, she might reply that the reflective, self-aware activity she has in mind is not simply a self-directed intentional action, but rather is a special, *sui generis*, sort of activity, one which provides us with the control over ourselves required for responsibility by allowing us awareness of and discretion over ourselves.

---

[9] The appeal to reflection extends far beyond Frankfurt. Just two more examples: Korsgaard famously connects the capacity to "step back" and bring one's perceptions and instincts "into view" with the capacity (and need) to believe or act for reasons. When we bring these features of our mind "into view," they no longer "dominate" us, and so we gain a kind of freedom over ourselves—but a freedom that requires us to act on reasons. See Christine M. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996). 92–93. (Though Korsgaard's views on responsibility are subtle, it seems that this kind of freedom is required for it.) T. M. Scanlon claims that we are responsible for our judgment-sensitive attitudes—attitudes that change (insofar as we are rational) in response to our judgments about their justification—that is, they change in response to reflective judgements about them. See T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998). (especially chapters one and six).

In reply, I will grant that there may be such *sui generis* reflective activity and that it may be important for many things.[10] However, it seems to me that we are owed some account both of what this activity is and, crucially, why it, with whatever features it boasts, does the job of grounding or conditioning our responsibility—whatever *that* is.[11]

3. My Alternative, as Illustrating the Hoped for Explanatory Connection

To illustrate the lack, I will begin by sketching the account of responsibility I favor. To be *responsible* for something, as I will understand it, is to be open to certain sorts of assessment on account of that thing, and, depending on the outcome of that assessment, to be the appropriate target of certain sorts of reactions on account of it.[12] Again, we can be responsible for our intentional actions, if we can be responsible for anything: we can be, on account of our intentional actions,

---

[10] E.g., the capacity to think about one's own thoughts is doubtless required for what Tyler Burge calls critical reasoning. See, e.g., Tyler Burge, "Our Entitlement to Self-Knowledge," *Proceedings of the Aristotelian Society* 96(1996). (Note that, even in this article, "rational control" is characterized as the ability to think about and "alter" one's thoughts, from the "point of view" of higher-order thoughts—the model recalls ordinary action.) Taking a more radically different approach, Matthew Boyle suggests that to believe at all is to be tacitly aware of your beliefs—that beliefs, and other states of mind, are activities such that, to partake in them is also to be aware of them. See Matthew Boyle, "Transparent Self-Knowledge," *Proceedings of the Aristotelian Society* Supplementary Volume 85(2011). What we lack, I contend, is an account of the relation between such sui generis reflection or reflective activity and responsibility.

[11] So, one might say that the changing of judgment-sensitive attitudes under critical reflection is not itself an action, but rather simply an aspect of the well-functioning of one's rational capacity. And, one then adds, the capacity for this reflective use of one's rational capacity is required for responsibility. My question is, why? In particular, why is this higher-order, reflective sensitivity better suited to secure responsibility than the capacity to form the attitude, or to make the judgement about reasons, itself? (Once we grant that there is a *sui generis* sort of activity that grounds or conditions our fundamental responsibility for our intentional actions, it is no longer clear why that activity should share the familiar features of those actions: if this reflective activity is not a kind of self-directed action, why expect it to involve discretion and awareness?)

[12] While I take this to be a kind of definition or account of what it is to be responsible, it is enough for the argument if one grants the biconditional: "*x* is responsible for *y* just in case *x* is open . . ." The account owes much to T. M. Scanlon's "Responsibility," in his Scanlon, *What We Owe*. (Scanlon's account, in turn, owes much to Peter F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* xlviii(1962).) This account of responsibility is distinguishable from another, closely-related usage, according to which to be responsible for a thing is to have obligations with respect to that thing.

One might ask whether I mean to be giving a "normative" or a "descriptive" account of responsibility: whether I mean to say that you are responsible just in case you are, as a matter of cultural fact, open to certain sorts of assessments and (taken to be) the appropriate target of certain reactions, or rather just in case you are *rightly* open to such assessments, and so the appropriate target of certain reactions. I mean the latter, though I take it that what a person is rightly open to assessment, etc., for will depend in complicated ways on contingent facts of culture, including facts about whether that person is, as a matter of cultural fact, taken to be open to assessment.

open to assessment not only as reasonable or unreasonable, justified or unjustified, but also as greedy, gracious, petty, courageous, magnanimous, insensitive, and the like. If one is responsible then, in light of such assessments, one can be the appropriate target of certain sorts of reactions, such as resentment, gratitude, admiration, trust, distrust, or esteem.[13]

Notice that we can also be responsible for a wide range of things other than our own intentional actions. We can be responsible, in the sense suggested, for the misbehavior of our dog, the disarray of our apartment, the operation of our digestive system, or the functioning of our automobile. We can be open to assessment on account of the misbehavior of our dog or the failure of our worn-out brakes, and, depending on the outcome of that assessment, we may be the appropriate target of the relevant sorts of reactions. We might be thought careless, negligent, indulgent, or sentimental; we might be the object of resentment, indignation, outrage, or distrust.

Plausibly, the responsibility we bear for this latter range of things is explained, in part, by our responsibility for our intentional actions. What responsibility you bear for your dog's behavior or the functioning of your brakes derives from the fact that these are things you can affect and so perhaps control through your intentional actions, together with the fact that they somehow fall into your jurisdiction—that is, together with the fact that you are rightly expected to affect and control them in certain ways. So, e.g., you have obligations with respect to your dog and your car, and if you

---

[13] Many many of these reactions seem to me to contain, or presuppose, or imply, or evoke an evaluation of the kind just mentioned—so that the evaluation and response need not unfold in two wooden stages.

There are also assessments of and reactions to a person on account of things for which that person is not (necessarily) responsible—being beautiful and therefore admired, highly contagious and therefore avoided, etc. I will not attempt the difficult task of specifying the range of assessments and reactions associated with responsibility; I trust the reader can locate the central cases.

It is also worth mentioning that, on the current definition, one is responsible for things that are morally innocuous (intentionally dropping one's keys on the desk, choosing vanilla rather than chocolate ice-cream). One is *open* to assessment on account of them; the outcome of the assessment would be neutral, and no particular reactions would be warranted.

(I leave aside the question of sanctions and punishment. Some will think the justification of these follows more-or-less directly from the fact that one is responsible, others will think they are subject to further, and different, standards of justification. For present purposes, we need not settle this dispute.)

neglect these obligations you will be criticized for it.[14] While you do not (it seems to me) have obligations with respect to the disarray of your apartment, we nonetheless rightly think of your apartment as yours to manage, and so take its state to reflect upon you. So I will say you are responsible for such things because they fall into your *jurisdiction*: you can affect and control these things through your intentional actions; they are, in some sense, yours; and so you are open to assessment on account of them.

Note that jurisdictional responsibility presupposes responsibility for our intentional actions. Thus, we are not—and crucially, we could not be—responsible for our intentional actions simply because they fall into our jurisdiction. That is, we cannot explain our responsibility for our intentional actions simply by appeal to the fact that they things that we are rightly expected to affect and control through our intentional actions. To think so would launch an immediate and vicious explanatory regress.[15] Rather, if we can be responsible for things because we can affect and control

---

[14] The neglect need not be intentional. To say that you are open to assessment on account of your dog's behavior or the state of your apartment is to say something more than that you are open to assessment on account of the actions by which you have affected your dog or your apartment, or on account of the decisions you made to neglect your dog or your apartment. It is to say that you are open to assessment on account of your dog's behavior or the state of your apartment, even when these bear no immediate relation to any particular action you took or particular decision you made. Rather, you are open to assessment because they are in your purview, and so speak about you—including, perhaps, your concerns, priorities, and patterns of attention and inattention. For discussion of our responsibility for patterns of attention and neglect see Angela M. Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115, no. January (2005).

[15] The threat of regress has been a persistent source of a certain form of skepticism about moral responsibility. See, e.g., what Galen Strawson calls "the Basic Argument" in Galen Strawson, "The impossibility of moral responsibility," *Philosophical Studies* 75(1994).

them through our actions, our responsibility for and control over our actions must be explained in some other way.[16]

3.1. ANSWERABILTY

I would suggest that we elaborate and explain this more fundamental sort of responsibility by considering what I will call *answerability*, a notion I take, roughly, from Anscombe.[17] Anscombe notes that, whenever one intentionally ϕ's (where ϕ stands for some ordinary action, such as doing the dishes or dismissing the students), one can rightly be asked, "Why are you ϕ-ing?," where this question looks for what she calls a "reason for acting."[18] Such a why-question is, in Anscombe's terms, "given application" whenever one acts intentionally. Drawing on her insight, I will say that one is *answerable* for one's intentional actions, where one is *answerable* just in case a request for one's reasons is given application.[19]

This notion of answerability is somewhat subtle, in part because the notion of a reason for acting is difficult and in part because the sense in which the question is rightly asked, or "given application," is not obvious.

Consider, first, *reasons for acting*. Sometimes, when thinking about reasons for action, philosophers have in mind those psychological states or events (typically beliefs and desires) that would explain the action. Other times they have in mind those (typically non-psychological) facts

---

[16] Steven Gross points out, in conversation, that one might think that our responsibility for each intentional action is secured by the fact that we could affect and control it taking another intentional action. This might answer a worry about how a certain condition or requirement on responsible action is met (if we can be responsible for a thing only if we can act upon it, then we can be responsible for our actions because we can act upon them by taking other actions). (Galen Strawson has in mind a different condition. If I understand him correctly, he thinks you are responsible for a thing only if you brought that thing about by a responsible choice, and that, accordingly, you are responsible for a choice only if you brought that choice about by a prior, responsible choice. This would, I think, launch a regress.) However, while appealing to the fact that we can act upon our own actions might allow us to satisfy a condition or requirement for responsibility, it will not, I think, help us to understand *why* we are responsible for our intentional actions and their effects. This latter, explanatory question is the one that concerns me.

[17] G. E. M. Anscombe, *Intention* (Oxford: Blackwell Publishing Co., 1957).

[18] Ibid., 9.

[19] My account of answerability draws on, but is not wholly faithful to, Anscombe's.

that would justify the action. Anscombe's why-question asks for neither. I will understand it to ask for (what I will call) *the agent's reasons for acting*, that is, those considerations (i.e., those facts or purported facts) that the agent took to count in favor of acting, the so taking of which (in part) explains the action.[20] Suppose you left the conference early. I might ask you why, and you might answer, "Because there was nothing else worth seeing." We can usually assume that this consideration, *that there was nothing else worth seeing*, was (among) your reason(s) for leaving.[21] This reason is not, itself, a mental state, and it may not justify your action. Indeed, it might not even be true. Nonetheless, this consideration plays a role in explaining your leaving, insofar as you took it to count in favor of leaving, decided (partly) on account of it to leave, and so left. In asking the Anscombean question, one is asking for reasons which play this sort of role: considerations the agent took to count in favor of so acting, the so taking of which will (in part) explain the action. That they play this role in explaining the action makes them the agent's reasons for acting.

Consider, next, when a why-question that looks for the agent's reasons for acting is *given application*, or *rightly asked*. On a natural reading, a question is rightly asked just in case the questioner is justified in posing the question. This is not the sense of 'rightly asked' at issue. Whether the questioner is justified in posing a question depends, in large part, on facts about the questioner: what she knows, what assumptions she is justified in making, what obligations she is under, etc. In the sense presently at issue, whether the question is rightly asked depends instead on facts about the one questioned—on facts that show that she is answerable.[22]

---

[20] These are what Scanlon calls "operative reasons." See Scanlon, *What We Owe*: 19. I consider the relevant sort of explanation, and compare it to Davidson's more familiar account, in Pamela Hieronymi, "Reasons for Action," *Proceedings of the Aristotelian Society* 111(2011). My account of the agent's reasons for acting is not wholly amenable to Anscombe's way of thinking. She tends to provide, as reasons, descriptions of the larger action in which one is engaged. However, she does believe that we must be able to find what she calls a "desirability characterization," and this will bring her account close to the one I offer.

[21] In assuming this was among your reasons, we are assuming not only that you have answered sincerely, but also that you are correct about your own reasons—you might be sincere but mistaken.

[22] I am grateful to Mark Greenberg for help with this clarification.

Consider, then, Anscombe's own reflections on when her question is "given application." As noted, Anscombe thinks the request is given application by intentional actions—importantly, she thinks it is given application even by an intentional action that was not done for any particular reason. She says, "the question ['Why are you ϕ-ing?'] is not refused application because the answer to it says that there is *no* reason, any more than the question how much money I have in my pocket is refused application by the answer 'None'."[23] Rather, according to Anscombe, the question is refused application by the answer, "I didn't know I was ϕ-ing," just as (presumably) the question "How much money do you have in your pocket?" would be refused by the answer "I have no pockets." The latter question is refused because an assumption made in asking it—that you have a pocket—is shown false. It seems, then, that a question is given application just in case the assumptions naturally made in asking it are met.

What, then, is the assumption that gives application to a request for one's reasons? It cannot be the assumption that one *has a reason* for ϕ-ing: that assumption would be false, and so the question refused, if one ϕ-ed for no reason.[24] I suggest it is rather the assumption that the person has, in some sense, settled for him or herself (positively) the question of whether to ϕ. It should be uncontroversial that to intentionally ϕ for certain reasons is to have, *in some sense*, settled the question of whether to ϕ, for those reasons.[25] The Anscombean question inquires after the reasons, if any, that you take to bear on this question.[26] The reasons the why-questions looks for, retrospectively, are just the reasons for which one would, prospectively, settle the question of whether to act. Quite

---

[23] Anscombe, *Intention*: §25.

[24] Some would have it that we cannot act without reason. I think we can avoid this contentious claim, while securing its benefits, by making the claim I am about to suggest: to intentionally ϕ is to have settled for oneself the question of whether to ϕ—a question on which reasons can bear, but which one might settle for no particular reason.

[25] As will become clear, the appeal to "settling a question" is not meant to introduce an additional psychological state or event. Rather, the claim "to intend to ϕ is to settle positively the question of whether to ϕ" simply notes the uncontroversial conceptual connection between an intention and a positive answer to the question of whether to ϕ.

[26] Anscombe herself allows a slightly wider class. I would hope to argue that the exceptions prove the rule.

generally, if you have settled a question for yourself, it seems that you can rightly be asked for the reasons, if any, that you took to settle it.[27] It seems, then, that the question "Why are you ϕ-ing?" (or, the un-Anscombean question, "Why did you ϕ?") asked of a particular person, is given application by the truth of the assumption that the person is ϕ-ing (or ϕ-ed) intentionally—that is, that the person is ϕ-ing (or ϕ-ed) because he or she has settled the question of whether to ϕ.[28] So I suggest that, whenever one intentionally ϕ's, one, in some sense, settles for oneself the question of whether to ϕ, and that this settling grounds and explains one's answerability.

3.2. Answerability as the Fundamental Form of Responsibility

This account of answerability will also, I believe, ground and explain our responsibility for intentional actions (in the sense of responsibility sketched above).

Note that, in revealing your positive answer to the question of whether to ϕ, your intentional actions therein reveal something of your mind. Your answer to this question will cohere, more or less imperfectly, with other things that you believe and intend, and so reveal a certain stretch of your mind—it will reveal what you find worth doing, and, by extension, something of what you think

---

[27] Some will think that that one can be asked for reasons whenever one has settled a question, because, if one has settled a question, then one *should* have had reasons for doing so, and, and request for one's reasons is given application if and only if one ought to have had reasons. (I owe this suggestion to Mark Greenberg.) While this thought is helpful, in drawing attention away from the questioner, I would modify it somewhat, in order to leave open the possibility that one can sometimes settle a question for no particular reason, without criticism. Rather than claim that, whenever one has settled a question, one *should* have had reasons for having done so, I would say something weaker: settling a question is the kind of thing that can be done for reasons—there is, so to speak, a place for one's reasons, or reasons would be apt. This weaker claim seems to me sufficient to give application to the request for reasons, and it seems to me to be what Anscombe was pointing out, in making her "grammatical" point.

[28] The question is likewise readily refused application (as Anscombe claims it should be) by the claim that one did not know that one was ϕ-ing: absent the possibility of unconscious decisions, the claim that one did not know one was ϕ-ing undermines the assumption that one is ϕ-ing because one settled the question of whether to ϕ. See Ibid. For an interesting discussion of Anscombe's claims, see Kieran Setiya, "Explaining Action," *The Philosophical Review* 112, no. 3 (2003). (The last two paragraphs repeat ideas, and sometimes sentences, which also appear in Pamela Hieronymi, "Responsibility for Believing," *Synthese* 161, no. 3 (2008); Pamela Hieronymi, "The Will as Reason," *Philosophical Perspectives* 23(2009).)

true or valuable. If we further know something about the reasons (if any) for your positive answer, then we will know something more of your mind. We may form an idea of the quality of your will.

Suppose, for example, that you intentionally end the fight. We know, then, that you settled for yourself (positively) the question of whether to end it. If we know a little about the context of the fight, and a little bit about your particular epistemic situation, knowing that you decided to end the fight tells us something of how you think about the world and your place in it. We will react in ways that reveal that we find your decision reasonable or unreasonable, justified or unjustified. If we further think you decided to end it for certain more-or-less-elaborated reasons, we may form certain further, more-or-less elaborate opinions about you: we might think you have been disloyal, spineless, magnanimous, mature, or conniving. Such assessments are typically thought to license certain corresponding sorts of reactions: resentment, contempt, regard, admiration, or distrust.[29]

So it seems both that one is answerable for ϕ-ing just in case one has settled for oneself the question of whether to ϕ and that settling that question generally leaves one open to the sorts of assessments and reactions, openness to which amounts to being responsible for ϕ-ing (at least in the sense here sketched). And so it seems that one will be responsible for ϕ-ing whenever one is answerable for ϕ-ing, for the same underlying reason.

Thus the claim that acting intentionally involves settling for oneself a question allows us to see, at least a little bit more clearly, how and why we are responsible for our intentional actions. Earlier we saw that we could not say that we are responsible for them simply because they fall into our jurisdiction—because we can control and affect our intentional actions through our intentional actions. Rather, there must be some more fundamental or original way in which we are responsible for our intentional actions. We can now say that we are responsible for our intentional actions because they reveal our answering of a particular question about a particular action in a particular

[29] Other things being equal—that is, absent certain familiar excuses—the assessments and reactions we will have are just those one is open to when responsible The role of excuses is important and difficult.

context, and so reveal something of our mind, or self. But this mind or self just is the object of the relevant sort of assessment and reaction, when one is responsible.[30]

3.3. Settling A Question as Explaining Responsibility

If this is right, then there is a natural alternative to the reflective account. Whereas the reflective account models the fundamental activity that grounds and explains our responsibility as a kind of self-directed action (or as a *sui generis* activity that shares the features of action), I am suggesting that we model the fundamental activity in a different, but also, I think, natural, way: as the settling of a question. Settling a question seems a lot like making a decision or a choice, and if anything were an uncontroversial locus of responsible activity, it would be decision or choice. Indeed, it seems natural to solve our original puzzle about why we are responsible for our intentional actions by claiming that we are responsible for them, not because they fall into our jurisdiction—not because we can affect and control them through our intentional actions—but rather because they reflect our decisions or choices.

3.4. The Unorthodox Road

However, in adopting this model, and, in particular, in adopting it for the reasons here given, one takes a fateful step down a perhaps unorthodox road. Notice that the Anscombean idea of answerability will naturally—practically effortlessly—extend far beyond the case of intentional action, as will the sort of responsibility of which it seems to be a species. Most obviously, you are answerable, not just for your intentional actions, but also for your *intentions*. If you intend(ed) to ϕ, then—whether or not you actually ϕ—you can rightly be asked, "why do you (or did you) *intend* to ϕ?" where this question looks for your reasons for ϕ-ing. This answerability is easily enough accounted for by the uncontroversial claim that one intends to ϕ only if one, in some sense, settles

[30] The forgoing argument applies to intention the argument made for belief in Hieronymi, "Responsibility for Believing."

for oneself the question of whether to ϕ. Settling that question, one is answerable for the reasons, if any, one takes to bear positively on it—again, this seems a lot like making a choice or decision, whether explicit or merely tacit.

But as easily as this extension to intention is made, to make it is to take a fateful step. The kind of agency or activity we here take for granted—the agency we exercise with respect to our *intentions* when deciding what to do—differs significantly from the control we exercise over our *actions*, when deciding what to do. In particular, our agency with respect to our intentions lacks both discretion and awareness. I will explain:

We have already noted that, when we act intentionally, we enjoy both a certain kind of awareness of and a certain kind of discretion over our actions. Indeed, we can now see that the model of settling a question readily explains these facts: if we act intentionally by settling for ourselves a question that represents the action under some description, then we, in some sense, have in mind what we are doing: we are in some sense aware of what we mean to be doing, because the relevant question includes a representation of the action, under some description. And, because we can, generally, settle any question for any reason we take to bear sufficiently on it, we can decide to do that which we represent for any reason we take to bear sufficiently upon the question of whether to do it. You can, e..g, raise your right hand, or turn off the music, or say something mean, in order to

win a bet, make a joke, relieve your boredom, or please your partner. We thereby enjoy discretion: we can decide to act for any reason we take to show the action worth doing.[31]

Notice, though, that, perhaps surprisingly, we do not—in fact, we could not—enjoy the same sort of discretion with respect to our intentions. You cannot decide to *intend* for any reason you take to show intending sufficiently worth doing. You can only intend for reasons that you take to show the *action* sufficiently worth doing.

This claim sometimes meets resistance. Gregory Kavka's toxin puzzle provides a case in which you might take yourself to have reason to intend but not to act.[32] But more mundane cases will do. Suppose, e.g., that you have no intention of marrying your partner, and that he or she is unhappy about this fact. And suppose your partner cares far more about your state of mind than about the legal arrangement. Suppose, further, that you are generally eager to please your partner, and, indeed, you would be quite willing to house the intention to marry, so long as you did not have to actually endure the wedding or enter the legal relationship. In such a case, you may take yourself to have sufficient reason to intend to marry your partner, though not sufficient reason to go through with the wedding. But in such a case you cannot form the desired intention. You will intend only if you

---

[31] Two features of intentional action are often commented upon: that actions are intentional only under some description(s) and that, when one acts intentionally, one should somehow know, "without observation," what one means to be doing—one should know what one means to be doing in a way that one does not know, without observation, either the unforeseen consequences of one's actions or the true descriptions under which one's action was unintentional. (Anscombe, *Intention.* is perhaps the classic statement of the claim that we know "without observation" what we are doing. I find a useful statement of the basic philosophical puzzle in Keith S. Donnellan, "Knowing What I Am Doing," *The Journal of Philosophy* 60, no. 14 (1963).

Both features can be accounted for, on the present proposal, by appeal to the fact that, in acting intentionally, one has settled for oneself the question of whether to do that—one has settled for oneself a question that represents the action under some description. The action is intentional only under that description, and it is natural to think that, having settled that question, one knows, in some sense, without observation, what ones means to be doing. On this picture, to act intentionally is to be, in some sense, the cause of one's own representations—the cause of that which one represents to oneself, in settling the question of whether so to act.

The sense of representation here is obscure (as is the sense in which we are "aware" of what we mean to be doing). In fact, I think to talk of "representations" can be misleading, insofar as it may lead one to look for distinct psychological states that do the representing—distinct ways in which one is aware of what one is doing. See, e.g., the papers collected in Johannes Roessler and Naomi Eilan, eds., *Agency and Self-Awareness* (Oxford: Oxford University Press, 2003). The present claim to seems to me of a different order—though not completely isolated from such investigations.

[32] See Gregory Kavka, "The Toxin Puzzle," *Analysis* 43(1983).

are committed to act. Thus it seems we do not exercise the sort of control over our intentions that we exercise over our actions—we cannot form, revise, or maintain them for any reason we think shows *that* worth doing.[33]

Note, further, that, if the account I have given of our fundamental responsibility is correct, fundamental responsibility actually *requires* such a lack of discretion. We are fundamentally responsible for a thing, we said, because it reveals our take on the world and our place within it—it reveals what we find true or valuable or important. But we *cannot* enjoy discretion with respect to whether we find something true or valuable or important—we cannot enjoy discretion over takings or findings true or important. You might, e.g., think that the possibility of winning a bet or making a joke provides you with very good reason to take something to be true. But if you represent something as true, for these reasons, that representation will not reveal your take on what is true.[34]

---

[33] Niko Kolodny points out that, in any such example, any reason against acting will also be a reason against intending so to act, since your intentions are likely to lead to action. Because ϕ-ing is an obvious consequence of intending to ϕ, Kolodny doubts that there are any cases in which one has sufficient reason to intend to ϕ but lacks sufficient reason to ϕ. Perhaps, then, you *can* intend to ϕ for any reason that counts sufficiently in favor of so doing. It just turns out that you will have such reasons only in cases in which you also have sufficient reason to ϕ.

But even if one established that the only considerations that in fact count sufficiently in favor of intending are those that count sufficiently in favor of acting, and so established that a person can rightly intend to ϕ for any reason that (in fact) counts sufficiently in favor of so doing, one would not thereby undermine my claim. My claim is that, while you can (intend to act, and, providing all goes well) act for any reason that you take to count sufficiently in favor of so acting, you cannot intend to ϕ for any reason that *you take to* count sufficiently in favor of doing intending. So, to undermine my claim, one would have to establish, not just that the only reasons for intending are those that are (in fact) reasons for acting, but that no one could *take* reasons to count sufficiently in favor of intending without also *taking* them to count sufficiently in favor of acting (Nishi Shah is aiming at something like this position, with respect to belief, in his Nishi Shah, "How Truth Governs Belief," *The Philosophical Review* 112(2003).). But it seems possible that someone might take that view, even if it is mistaken. So, suppose someone (perhaps mistakenly) thought that his partner's unhappiness was reason enough to intend to marry, without taking it to count sufficiently in favor of marrying. My claim is that such a person cannot intend for the reasons that he takes to count sufficiently in favor of intending, though he could act for *any* reason that he takes to count sufficiently in favor of acting.

To put the point another way: you will intend to ϕ only if you are committed to ϕ-ing, and (if you commit to ϕ-ing for reasons) you can only commit to ϕ-ing for reasons that you take to settle the question of whether to ϕ. But you might (perhaps mistakenly) take certain considerations to show intending to ϕ worth doing, which you do not take to show ϕ-ing worth doing. You will not be able to intend for these reasons (though, as noted, you may be able to bring it about that you intend for those reasons). In contrast, you can (intend to ϕ and, providing all goes well) ϕ for any reason you take to show ϕ-ing worth doing.

[34] I have found instructive J. David Velleman, "On the Aim of Belief," in *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).

It will rather reveal your take on what is worth doing—viz., representing this as true, in order to win a bet or make a joke. Likewise, if you represent some action (getting married or drinking the toxin) as to be done for some reason that you take only to show it good to *represent it* in that way, that representation does not show that you take the *action* as to be done—it would, instead, show that you take, as to be done, *representing* that action as to be done.

So, in general, we cannot find something true, or valuable, or important, for any reason that we thinks shows *finding it so* worth doing—because responding to certain such reasons will not qualify as finding the relevant thing true or valuable or important. Thus, we cannot enjoy discretion of the sort described over what we find true or important or worth doing. So, if we are, most fundamentally, responsible for our take on what is true or important or worth doing—if our take on these questions is the object of the assessments and reactions that are characteristic of holding someone responsible—then we *cannot* enjoy discretion with respect to those things for which we are most fundamentally responsible.[35] This extremely important point is too often overlooked. (In fact, people often enough assert that we can be responsible only for what we do voluntarily, and mean, by voluntary, what I mean, here, by discretionary. This is false.)

Turning, now, to awareness: Notice that, without the ability to effect whatever change you find worth effecting, the importance of awareness is far less clear. It is unclear why it helps to be *aware of* having a take on certain objects, if that awareness will not provide you with discretion over your take. (Think, again, of Kant's creature.)

Of course, if you are aware of your mind, you may be able to take action to change it. You might be able to manage or manipulate your own thoughts. But this is just the mundane way in which awareness of anything will enhance your control over it. If I remain aware of the whereabouts of my dog or my child, I will be in a better position to control him or her. We were

---

[35] This is a point I make at length in Hieronymi, "Responsibility for Believing.", where "voluntary" stands in for "discretionary." A restatement of my earlier argument that we cannot adopt these attitudes "at will" appears in Pamela Hieronymi, "Believing at Will," *Canadian Journal of Philosophy, Supplementary Volume* 35(2009).

wondering, instead, how awareness might enhance, not the control you exercise over your mind by taking action to affect it, but rather the control you exercise over your mind by settling questions about whether something is the case or whether to act in some way. We were wondering whether lacking an awareness of your mind while forming or holding an intention or belief leaves one with less control over the forming or holding of it, or whether having such an awareness would increase one's control. I do not see how—once we set aside the ways in which awareness enhances self-management.[36]

Thus it seems the agency we exercise with respect to our intentions, when deciding what to do, lacks the two features that provide us with our most familiar sense of control: discretion and awareness.[37] Admittedly, it can seem very peculiar to think that we can be exercising agency with respect to something of which we are not aware and which we did not intend—over which we do not enjoy discretion. It seems to me, though, that we should simply accept that certain forms of agency do not sport these features. The first reason is that, argued above, intending in fact lacks these features, and yet we must exercise agency or be active when intending, if we are to be agents when acting. If we are not agents when making our decisions, it is hard to see how we could be so, in executing them.[38] The second reasons is that, again as argued above, if we are responsible, most

---

[36] Whether or not awareness would help, I doubt that we in fact have our intentions in mind in anything like the way that we have in mind what we intend to do. It seems that you are typically aware of your intentional actions in that you know, in some sense without observation, what you mean to be doing—in a way that you do not know either the true descriptions under which your action is unintentional or the unforeseen consequences of your action. I doubt we know, in this same way, that we intend. "Creating an intention" need not be any part of the description under which one's action is intentional, and, in a suitably constructed case, one's intention might be an unforeseen consequence of one's action. (Perhaps there is some other way in which we must know of our own minds. See, e.g., Boyle, "Transparent Self-Knowledge.")

[37] These claims mask a considerable amount of subtlety, and establishing them requires a good deal of work. I argue that one cannot "intend at will" in both Pamela Hieronymi, "Controlling Attitudes," *Pacific Philosophical Quarterly* 87, no. 1 (2006). and Hieronymi, "Believing at Will."

[38] I am sometimes asked why this thought does not prove too much: presumably our intending is itself a product of things that are no our activities, and yet I claim we are active when we intend. So why not think that we are not active when we intend, but only when we execute intentions in action? I will only say that I cannot make sense of a picture in which my intentions are formed, passively, and and my role, or my active role, lies only in executing them.

fundamentally, for our take on the world, then we *cannot* enjoy discretion over the ultimate objects of responsibility.

Thus, on the account of responsibility I have offered, we must lack discretion in exercising the fundamental, responsible agency with respect to our own minds. The importance of awareness seems questionable, as well. Once we abandon these features, reflection seems much less appealing as a ground or condition for responsible agency. If one wants to insist that a special sort of *sui generis* reflective activity both grounds and conditions responsible agency, one needs to explain not only what sort of activity one has in mind, but also how it explains or why it is required for responsible activity.

4. UPSHOT: A DUALITY OF RESPONSIBILITY

Before turning to the second reply of the champion of reflection, I want to note an important upshot of this way of modeling responsible activity: on this model, we are responsible for certain states of mind in two distinct ways.

To see this, consider again your dog. Earlier we noted that you are responsible for your dog's misbehavior because it falls into your jurisdiction. But you are not *answerable* for your dog's misbehavior. You cannot be asked for your reasons for his misbehavior—that makes no sense—because his misbehavior cannot be understood to have come about because you settled for yourself

the question of whether so to misbehave. You are responsible for the misbehavior, not because it embodies your answer to a question, but simply because it falls into your jurisdiction.[39]

It is extremely important to note that the things for which one is answerable can *also* fall into one's jurisdiction. For example, your beliefs are facts about you that you can affect and perhaps control through your intentional actions, and sometimes you are expected to do so. If you are assigned to a committee that awards a prize, concerns of neutrality might dictate that you not learn the identity of the candidates. You are expected, then, to avoid learning them. You might excuse yourself from certain conversations. If you do not, when it was clear you ought to have done so, you are responsible for your negligence. If you thus come to believe that candidate number four is Jones, you are responsible for your belief in two distinguishable ways. You are answerable in the usual way—you can be asked why you believe Jones is candidate four, where this question looks for reasons that you take to show that Jones is candidate four. But you are also responsible for the fact that you believe Jones is candidate four, in much the way that you are responsible for the misbehavior of your dog: you can sometimes control whether you believe, through your actions, and sometimes we expect you to do so. You failed to do so, in such a case, and you are therefore open

---

[39] You are answerable for a great many things that are neither attitudes nor actions, but rather are the intended or foreseen products or consequences of your intentional actions. I can be answerable for the illumination of the room or the alerting of the prowler—I can be asked why I illuminated the room or alerted the prowler—if either was the intended or foreseen outcome of my intentional action of, say, turning on the light. I can be assessed and responded to on account of either, if either obtains either because I settled for myself the question of whether to bring it about or because I settled for myself the question of whether to do something else, this foreseen consequence notwithstanding. (In claiming that one is answerable for the foreseen consequences of one's intentional actions, I part company with Anscombe (Anscombe, *Intention*: §25.).)

There are obvious difficulties with determining which of the intended consequences of an action should be included in the description of the action itself. If one turned on the light intending to alert the prowler, is one's action itself properly described as "alerting the prowler"? If so, that the prowler is alerted is not simply a state of affairs that is the foreseen consequence of your action; it is rather part of your action, itself.

The worry does not arise in the case of your dog's misbehavior. If you have trained your dog to misbehave on cue and you intentionally give the cue, then you are indeed answerable for the misbehavior, in the same way that you are answerable for the mess you made in the kitchen—you can be asked why you brought it about. That is, you are answerable for the *action* that created it. But it would be odd to think of either the mess or your dog's misbehavior as part of, rather than a consequence, of your action.

to the relevant sorts of assessment—as careless, unreliable, or perhaps unscrupulous—and to the corresponding range of reactions.[40]

The proposed account allows for this duality in our responsibility for and agency over our actions and attitudes. Nothing about embodying the answering of a question precludes an action or attitude from also being the target of another exercise of agency,[41] and so nothing prevents these actions or attitudes, understood as embodying our answers to questions, from also falling into our jurisdiction. Our responsibility for and agency with respect to the relevant actions and attitudes thus has two distinguishable aspects: we are *answerable* for them, insofar as they embody our answer to certain questions, and they fall into our *jurisdiction*, insofar as we are expected to manage and control them through our actions.[42] I take it to be a considerable strength of the account on offer that it can allow for, and indeed clarify, this duality.

5. Second Reply of the Champion of Reflection

I will now consider a second possible reply of the champion of reflection. To review: I claimed that, if there is a question about how or why we are responsible for our actions, we cannot answer it by appeal to a sophisticated, self-directed action. In reply, the champion claimed that she was not appealing to a self-directed action, but rather to a special, *sui generis* sort of activity. I then asked for an account of this activity and of its relevance to responsibility. I presented my own account of responsibility and corresponding account of responsible activity, to illustrate the lack.

---

[40] For an alternative example: if you are prone to outbursts of anger, you now have certain obligations of anger management. You are expected to avoid certain situations, e.g. If you do not, and you end up subjecting someone to an uncalled for but creative and colorful tirade, then you are can be assessed and responded to in two ways: for failing to avoid the situation and for the particulars of your tirade.

[41] The importance of this point should not be overlooked. It both allows us to manage ourselves in extremely important ways and allows for the (often objectionable) manipulation of another person's exercises of agency.

[42] Elsewhere I call the agency we exercise over our attitudes by settling a question or set of questions *evaluative control*, and the agency we exercise over our attitudes by taking actions designed to affect them *managerial* or *manipulative control*. These forms of agency are often exercised in tandem, and so often obscured by and confused with one another.

Rather than provide the requested account, the champion of reflection might pursue a different approach. She might adopt an account of responsibility, such as the one I have offered, which allows that we can be responsible for activities which do not, themselves, involve reflection, but she might insist, nonetheless, that a person cannot be responsible for such activities unless he or she is also capable of reflecting upon and changing his or her mind.[43] The champion here appeals to our *capacity* for reflection as a *condition* on responsibility. She claims that an exercise of our ordinary capacity to act—to think about and change things—will not qualify as responsible activity unless we are also capable of exercising that capacity reflexively: unless we can also think about and change ourselves. Our take on the world is not a take for which we are responsible unless we can also step back from, reflect upon, and change that take.[44] The question we must ask is, why should this be so? Why is responsibility subject to such a requirement?

I believe that the usual thought again concerns control: with the capacity to reflect comes the possibility of controlling or having controlled one's own mind in a way that includes discretion and awareness. It is hard to shake the thought that, unless one has the capacity to exercise control of

---

[43] The champion need not adopt my account. She could, instead, say that you are answerable for your attitudes, not because they embody your answer to a question—not because of any activity of yours—but simply because they are the kind of thing for which there *could* be reasons and they somehow belong to (or constitute) you. She would then add to this that you are not responsible for such attitudes until you have the capacity to reflect upon and change them. In the main text I am questioning this additional requirement, not the underlying account. The underlying account seems to me unsatisfying in the way Hobart's veiw seems unsatisfying: it does not require activity. You can be asked why you believe, intend, resent, trust, etc., even though these are states of mind that simply happened to you—that appeared in your psychology, absent of any activity of yours. (In the text I am explaining why adding the capacity for reflection does not add the right kind of activity.)

[44] A nearby reply of the champion of reflection would claim that reflection does not so much afford control as it does locate or identify the true or truly responsible self: I am responsible only for those aspects of myself that I endorse.

The question to ask, of such an account, is the one asked, in effect, by Gary Watson, long ago: why should I be identified with my reflecting self or that which that reflecting self endorses? (It is not a brute ethical fact—indeed, I think it no ethical fact at all—that I am responsible only for those parts of myself of which I approve.)

It is tempting to say that the reflective self is the responsible self, or is the true self, because it is the self that one has shaped through one's own activities. But it is not clear why my self-shaping activities should make me responsible, if I was not already responsible for the activities by which I did the self-shaping. And if we could give an account of how I am responsible for my self-shaping activities, it seems we would be done: we would not need to appeal to the fact that I have shaped myself. One might instead claim that that the true self is the one I *could* shape. One would then have returned to the position considered above.

this form over oneself, one is not a candidate for responsibility. But, again, why should this be? I believe the usual thought is as follows: once you have the *capacity* to reflect upon and change yourself, then there is a sense in which you *could have* exercised that capacity. (The sense of "could have" here need not be—and is probably not thought by the champion to be—incompatible with determinism. Rather, it comes with the notion of a capacity.) The capacity for reflection secures the possibility of self-management, which, in turn, secures a sense in which, in any given case, the responsible agent could have done otherwise.

5.1 Setting Aside Nearby Issues

While I find this a tempting position, I also find it obscure—and, in the end, dissatisfying. We need to keep asking: Why must it be the case that I could have reflected upon and changed myself, if I am to be responsible for what I did (or, for what I am now doing), in these sense sketched above—if I am to be subject to the evaluations and responses characteristic of responsibility?

There are three nearby points we must acknowledge and avoid confusing with the question at hand. The most important is this: the capacity for self-management is clearly required for what I have characterized as jurisdictional responsibility for yourself. If you could not think about your own beliefs, e.g., or could not take action to affect your beliefs, then your beliefs could not fall into your own jurisdiction: they could not be things that we expect you to manage. Thus, the fact that you have the capacity to reflect and act upon yourself explains, very neatly, why you can be expected to manage yourself—why, e.g., you can be charged with negligence or unscrupulousness for failing to excuse yourself from the conversation about Jones. We need to set this aside and focus on evaluations and responses that do not depend on the expectation that you self-manage.

To help clarify this point, consider an analogous case. Suppose I am impatient and tend to become irritated with the stumbling best-efforts of those less gifted than I, treating their simple inability as willful obstinacy, even though I know this is incorrect. I am, if attentive, able to catch or anticipate my error and correct myself. On some occasion, I fail to do so, and I express my

impatience. I believe I am now guilty of two failures: of the underlying impatience, and, additionally, of negligence (or lack of conscientiousness or attentiveness) in failing to manage it (and its expression). We can grant that I can be charged with *negligence* on account of my impatience only if (and because) I could have engaged in some self-management (e.g., I could not be charged with negligence if I had no earlier evidence of my own impatience or its inappropriateness). We want to set this aside and ask about my responsibility for the underlying impatience, itself. We are asking: why would my responsibility for my impatience, itself, depend on the fact that I could have better managed it?

There is a second nearby point we can also acknowledge and set aside: the possibility of reflecting upon one's attitudes is required, not only for self-management, but also for critical reasoning—for thinking about and calling into question your own thoughts and assumptions, understood as such. And so the possibility of critically reflecting on one's attitudes is required for what I would call *authenticity*. Authenticity is, very roughly, the good aimed at by liberal arts education. It requires the ability to step away from, think about, doubt, and re-evaluate both the assumptions that have shaped you and those shaping the world you now inhabit, in order to become your own person, even in the midst of it all. It requires a capacity for critical thinking, together with a kind of honesty and strength of ego.[45] Someone who has gained in authenticity is, indeed, more truly her own self.[46] She is also, in an important sense, liberated; she enjoys an important form of freedom. It is a kind of freedom philosophers should reflect upon and defend. However, *this* form of freedom is not required for responsibility. The inauthentic are nonetheless responsible.[47]

---

[45] I gesture more grandly toward this notion in Pamela Hieronymi, "Making a Difference," *Social Theory and Practice* 37, no. 1 (2011).

[46] The phrase cries out for philosophical investigation. I trust the reader to track, roughly, what it often meant by it.

[47] The Betas in Huxley's *Brave New World* lack this form of freedom (as do any number of ordinary adults)—and, moreover, they have been seriously wronged by those who have ensured that they lack this freedom—but they are not, for that, less answerable, nor are they exempted from moral and interpersonal demands and expectations.

Finally, we can acknowledge that, unless you are capable of thinking about your own states of mind, you cannot sensibly be asked your reasons for some state of mind—because you would not be able to understand the question. And so it seems you could not be answerable. One might try to conclude that you could not be responsible, in the sense sketched—you could not be open to the sorts of evaluation and responses characteristic of responsibility.

But the last step in this line of reasoning is incorrect. I have claimed that the fact that one has settled a question will explain both answerability and responsibility, but I have not claimed that answerability is required for responsibility.[48] The assessments and reactions in question are assessments of and reactions to the quality of one's will. And, if there were creatures (there might be) who could settle questions about what is true, worthwhile, or important, and who could have a take on the world and of their place in it, but who could not think about their own states of mind, as such, it seems that such creatures would nonetheless have a quality of will. We are asking whether it would be appropriate to assess and respond to *their* wills in the way characteristic of responsibility. The champion thinks it would not be, and, moreover, thinks it would not be *because* the creature lacks the ability to reflect upon and change its mind. We want to consider that claim. To do so, we need to notice that the position advanced by the champion is not the simple claim that, because a creature cannot understand a request for its reasons, and so is not answerable, that creature cannot be responsible.[49]

---

[48] (Though I have claimed that, if one is answerable, then one is responsible.)

[49] Three related points: First, what is needed to understand a request for one's reasons for one's own attitudes is not the ability to think about *and change* one's attitudes, but only the ability to think about them. Understanding the question requires a certain amount of cognitive sophistication, not a certain kind of reflexive or reflective agency.

Second, the psychological sophistication needed for answerability is not exactly the ability to think about one's own mind —because the request for one's reasons need not be couched in psychological terms. Rather than ask why you decided to go to the store, I can ask why you went to to the store. Rather than ask why you believe the butler did it, I can ask for evidence for his guilt—considerations that show him guilty. The cognitive sophistication required is the ability to think about reasons (which may turn out, for other reasons, to be inseparable from the ability to think about one's mind).

Finally, answerability does not require an actual exchange—I need not actually ask someone for her reasons before I rightly regard her as answerable, and I can so regard her even if she lacks certain abilities required in order to provide her reasons in response to my request. Perhaps we do not speak the same language. This would not make it the case that she must speak English before I rightly regard her as answerable.

So, we are asking why the capacity to think about and change your own attitudes is required for the fundamental form of responsibility we have isolated: why must I have some ability to manage my impatience, e.g., before it expresses a take on the world that is rightly subject to the sorts of assessments and reactions characteristic of responsibility—a take which others could regard as malicious or self-absorbed and rightly respond with attitudes such as resentment or distrust. We need to set aside the fact that the capacity to think about your own states of mind is required to understand questions about them, the fact that a capacity to think about and question your own states of mind is required for authenticity, and the fact that the capacity to think about and manage your own states of mind is required for jurisdictional responsibility for those states.

5.2 As a Condition on the Aptness of Reactive Attitudes

Perhaps the possibility of self-management is required for the aptness of the responses and reactions that typically characterize *moral* responsibility, in particular—reactions such as resentment, indignation, or gratitude. Perhaps *these* are not apt unless their target is capable of self-management. So, perhaps you cannot resent my evident disregard for your interests unless I could have better managed my decision—unless I could have brought it about, though some reflection, that I made a different decision. Or perhaps I am not a legitimate target of resentment on account of my impatience unless I could have avoided the impatience through some earlier or concurrent bit of self-management.

This seems tempting, but again we need to understand why it should be true: why would resentment for some piece of disregard be inapt in those cases in which its target could not have avoided the disregard or impatience through earlier or concurrent self-management?

One seemingly popular thought appeals to fairness. Being resented is undesirable, burdensome, something people generally want to avoid, and it is often not fair to impose a burden on someone unless he or she had some opportunity to avoid that burden, either now or in the past.[50]

While this is a powerful intuition, I think it is confused. I have tried to display the confusion elsewhere.[51] I will briefly attempt to give a sense of those arguments.

The idea that resentment can be unfair because its target did not have an adequate opportunity to avoid it gains its power in part by thinking of resentment (and other such responses) as a penalty, sanction, or punishment that we intentionally impose on wrongdoers. The fact that someone lacked an opportunity to avoid a given penalty often shows the penalty unfair. But to think of resenting as imposing a penalty is, I would argue, to confuse certain non-voluntary reactions, such as resentment, with intentional, punitive activities like guilt-tripping (trying to make someone feel bad about what they have done). Though the two often (unfortunately) co-occur, they are very different.[52] Once we set aside guilt-tripping and other intentional, punitive responses to wrongdoing, and focus, instead, on the non-voluntary changes in attitudes and relationships that typically accompany the negative actions and attitudes of responsible people, I believe the appeal to fairness looses much of its power.[53]

In fact, I have argued elsewhere that the unfairness imagined is the wrong kind of reason for criticizing an attitude like resentment: criticizing resentment because it is an unfair burden on the

---

[50] See, e.g., Gary Watson, "Two Faces of Responsibility," *Philosophical Topics* 24, no. 2 (1996). For the contrary position, see Scanlon, *What We Owe*: 282–90.

[51] See Pamela Hieronymi, "The Force and Fairness of Blame," *Philosophical Perspectives* 18, no. 1 (2004).

[52] While I would be happy to live in a world without guilt-tripping, I cannot imagine a world in which poor behavior—actions and attitudes which show disregard for the interests of others, say—did not meet with negative reactions. The fact that poor behavior will elicit some negative reaction in any society anything like our own is, I take it, a lynchpin of the argument in Strawson, "Freedom and Resentment.". I think his characterization of these attitudes as *reactive* has been under-appreciated. See Pamela Hieronymi, "Freedom, Resentment, and the Metaphysics of Morals," (in progress).

[53] Scanlon makes a similar argument against the appeal to fairness, without the emphasis on the voluntary, in Scanlon, *What We Owe*: 282–90.

one resented is like criticizing a belief by pointing out that it has bad consequences.[54] Without detailing that argument, I will here gesture towards an analogy.

Consider a less-freighted case: Being distrusted is burdensome, something people have an interest in avoiding. Yet the fact that someone lacked any earlier opportunity to make herself a more reliable person, and lacks the capacity, even now, to effect the required changes in herself, would not make your on-going distrust of her unfair. Your distrust simply marks the fact that she is unreliable; it is the way that fact manifests in your relationship with her. Her predicament, the fact that she cannot now improve herself and lacked any earlier opportunity to avoid her fate, may be tragic—perhaps even, in some cosmic sense, unfair. But the fact that she lacked any opportunity to avoid your distrust does not render your on-going distrust unfair. It is the wrong kind of reason to criticize your distrust.

So, too, I would argue that your resentment of my impatience marks the fact that you have been wronged by someone, the quality of whose will matters. It is the way that fact manifests in our relationship. The fact that I cannot now or could not earlier manage myself in such a way as to avoid the wrong may be tragic—perhaps even, in some cosmic sense, unfair. But the fact that your resentment is a burden that I could not have avoided does not render your resentment unfair. My lack of opportunity is the wrong kind of reason for criticizing your resentment.[55]

---

[54] See again, Hieronymi, "Force and Fairness." I have also given a principled account of what makes a reason for an attitude of the wrong kind in Pamela Hieronymi, "The Wrong Kind of Reason," *The Journal of Philosophy* 102, no. 9 (2005). I show the connection between reasons of the wrong kind and voluntariness in Hieronymi, "Believing at Will."

[55] Resentment admits of a distinction between the right and the wrong kind of reason. To see this, consider forgiveness. Forgiveness is often characterized as the forgoing of resentment. Resentment is burdensome not only to the one resented, but also to the one resenting. For this reason people are often counseled to forgiveness by appeal to their own self-interest: "Let go. You are only hurting yourself." But this counsel is about as effective as being told that you should believe everything will be okay because it will make you feel better in the interim. If you are lucky, you may be able somehow to take steps and manage yourself into the belief or out of your resentment, but you will not, in either case, have been given the right kind of reason for the desired change: you have not been given the kind of reason that would let you directly revise your belief or your resentment. (Compare being given evidence or a sincere apology.) So, you may instead be stuck in your resentment, unable to give it up, even though you see that you would be better off if you could do so. I discuss this is Pamela Hieronymi, "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, no. 3 (2001).

This view of resentment is sometimes resisted. People sometimes think of resentment, not as voluntary, but also not simply as a reaction to instances of disregard or disrespect, but rather as a reaction specifically to instances of disregard or disrespect that could have been avoided by somehow trying harder. Resentment is somehow understood to include, within it, a commitment to the claim, "and if you had paid attention and exerted yourself, you could have done better!"

Perhaps there is an attitude like this, and perhaps it is what many people call "resentment." I will call it "resentment-plus." Such an attitude will obviously require, on pain of some kind of incoherence or self-contradiction, a readiness to assent to the claim about self-exertion: one must be ready to believe that the one who is resented could have done better, by trying harder (either in the moment or at some point in the past). And it might seem that the truth of the claim about self-exertion requires that the one resented have the ability to reflect—that she is able to pay attention to herself and guide herself towards better behavior. So it may seem that the capacity for reflection is required before someone can be an apt target of resentment-plus, and, if being responsible requires being an apt target of this attitude, then it would seem that one could not be responsible unless one is capable of reflection.

It may be that intuitions about the need for a capacity for reflective agency are driven by concern about the aptness of resentment-plus. But I do not think we should allow this particular (and particularly unattractive) attitude to dictate our sense of what is required for moral responsibility.

Notice that, often enough, a person's sense of him or herself and his or her world would need to be thoroughly overhauled, before showing respect on some occasion could be a matter of simply paying more attention and/or exerting more effort. And, often enough, it is not the case that the person could be reasonably expected to have completed such an overhaul, prior to the offense. Consider, e.g., a case of entrenched chauvinism, in which the person's self-esteem, such as it is, depends heavily on the pride taken in being male. Place that person in a context in which that kind of chauvinism is widely accepted and in which he has not had his own attitudes remarked upon or

questioned. It seems right to say, of such a person, he could not have done better by paying attention and trying harder. Even so, such a person can show disrespect to others. And one could, I think, coherently and without self-contradiction, react in a such case with an attitude I would call "resentment" (perhaps others would like to call it "resentment-minus")—an attitude that (is neither voluntary nor punitive and) does not include any thought or commitment to the claim that the person could have done better by paying attention and trying harder, either on this occasion or through earlier efforts at self-improvement (any more than distrust includes a commitment to a claim about trying harder). The attitude instead responds, simply, to the disregard shown to one person by another. It is the negative reaction that marks that fact in one offended.

Sometimes, when thinking of the entrenched chauvinist confronting offense for the first time, people claim that the person offended should react to the chauvinist with a kind of faux-resentment, as (they say) one might react to a child. I find this implausible, paternalistic, and, I suspect, unnecessary. I suspect this recommendation is motivated by a desire to avoid resentment-plus—which is obviously out of place. But, rather than retreat to pretending-to-resent, we can retreat to resentment-minus. One might be genuinely offended—your interests and status have been disregarded by someone who matters in the way adults matter—absent any thought that the person could have done better by trying harder.

It might seem that, if we grant the possibility of attitudes such as resentment-minus, the game is over: we have granted the possibility of reactive attitudes that are not conditioned by a capacity for reflective agency, and it might seem we have granted that responsibility, in the sense we have been considering, is not so conditioned. But this would be too quick. The champion has another reply.

5.3 As a Condition on Moral Demands

Rather than consider whether the aptness of attitudes such as gratitude or resentment require that their target be capable of self-improvement, the champion of reflection might instead turn her attention to the sorts of demands whose violation resentment mark (moral demands, we might call

them) and claim that these *apply* only if the person who violates them had the capacity to satisfy them—either then and there or else through some earlier, possibly successful, project of self-improvement. Since we are not typically born able to satisfy the demands, and since even ordinary moral education does not bring us to virtue, the demands cannot apply unless we have a capacity to self-manage. But if the demands did not apply, they would not have been violated. And the fact that no demand has been violated surely renders resentment inapt.

I have also argued, elsewhere, against the first premise of this argument: that moral demands apply only to those who have the capacity to satisfy them, either then and there or through some earlier project of self-improvement.[56] (This thought is sometimes expressed in a slogan: "ought implies can.") I will again give a brief sense of the argument. The basic point is this: if moral demands were in this way custom-fit to each occasion, they would be both highly unusual among the demands we place upon one another and ill-suited for doing their job in adjudicating the interests of people needing to share a world peacefully.

Most demands that we face—the demands of parenting, say, or of being president—do not adjust themselves to the particular capacities or possibilities of those to whom they apply. Rather, we hope that the capacities of those who fill the role will expand to satisfy the unyielding demands. We hope that someone who is particularly self-absorbed, or particularly insensitive, will change in light of the needs of his or her children. But some vices—insensitivity is one—seem to guard against such improvement. Overcoming other vices requires psychological resources that not everyone has at hand. If the parent does not change, then he or she is condemned to be a bad parent and will suffer whatever consequences that entails. This may be tragic, but it is not (otherwise) inappropriate.[57] Many demands are similar.

---

[56] Pamela Hieronymi, "Rational Capacity as a Condition on Blame," *Philosophical Books* 48, no. 2 (2007). The same point is made in Pamela Hieronymi, "forgiveness, blame, reasons..." in *3am: magazine*, ed. Richard Marshall (2013).

[57] One might be tempted to think that it would be a kinder, gentler world if we would fit the demands to the person. I invite the reader to ruminate on such a world. I doubt it will prove attractive.

The exception are pedagogical demands. Pedagogical demands are custom-fit to the individual; they are rightly adapted to the particular, local abilities of the student—in fact, they are often rightly set just a tad beyond the current abilities of the student, so that, in practice, the student may—either by luck or by informed effort—happen upon the correct answer or movement or method, which he or she might then learn to repeat until he or she becomes proficient. But moral demands, I submit, are not pedagogical in this way.[58] I do not demand that you show just a little more concern for my interests than I think you are currently capable of showing, in the hope that you will eventually work your way to the fact that my interests matter as much as yours do.

Moral demands are, I believe, the demands placed upon us by our need to share a world peacefully with others who are (in Thomas Nagel's memorable phrase) equally real. And so moral demands are more like the demands of a hymn than the demands of an opera. An opera could be written to the ability of specific performers, and it could be revised if one of the performers on hand cannot mets its demands. A hymn, in contrast, must be written for the typical congregation. It should be written in such a way that most people can satisfy its demand tolerably well.[59] But there will predictably be some who cannot, and the hymn will not be re-written for them. They will simply perform poorly.

Of course, the demands of a hymn do not fall on those who cannot sing at all—whether well or poorly. And, I have acknowledged earlier, demands that one manage a thing cannot fall upon those without the capacity to manage it. The conclusion to draw, though, is not that demands must adjust until those to whom they apply are capable of *satisfying* the demand, but rather that the demands apply only to those able to partake in the activity in question, whether well or poorly. And so the analogous thing to say, when it comes to the basic moral demands, is that demands on the quality of

---

[58] If one thinks of moral demands as imposed by God, in a parental role, and then used by God, to judge one's fate, one will have two strong reasons to think otherwise.

[59] I believe this picture of the nature of moral demands underlies the argument in Strawson, "Freedom and Resentment." See Hieronymi, "Freedom, Resentment, and the Metaphysics of Morals."

one's will cannot fall on those who do not have a will of any quality—who do not have a take on the world and their place in it. But, for those who do, the fact that they could not, now or in the past, have had a will of better quality—the fact that they could not have avoided offense through some effort—does not show that they are exempted from expectations of regard and good will. In a slogan, vice does not exempt.

So, I would argue that the moral demands can stand, unyielding, in the face of an inability to meet them; that the reactions we have, when we regard one another as responsible, are neither punitive nor voluntary; and that these reactions need not include a commitment to the claim that their target could have avoided the wrongdoing by trying harder. If we grant these points, I believe it becomes very hard to see why it must be that I *could have* reflected upon and changed some attitude, either here and now or at some point in the past, if I am to be the apt target of the reactions characteristic of responsible agency. And so it seems hard to see why I must be capable of reflecting upon and changing an attitude before I am responsible for it.

6. A Place for Reflection?

It is, nonetheless, a striking fact that responsible creatures seem also to be creatures capable of self-reflection. If I hope to deny that reflection grounds responsibility by supplying control, or the ability to do otherwise, I must somehow account for this remarkable correlation between the capacity for reflection and responsibility. Here is a preliminary set of thoughts:

To be responsible, we have said, is to be the appropriate target of a certain range of reactions. Some of those reactions—resentment, indignation, gratitude, trust, betrayal—are characteristic of what we might think of as moral or interpersonal relationships. But *this* range of reactions (and the kind of relations constituted by them) is possible only for creatures capable of thinking, not only about another creature's *mind*, but also about another creature's *reasons*. I can resent you only if I can think about your reasons—only if I can, so to speak, contemplate your maxims. And, to think

about your reasons, it seems I need to be able to think about your mind—to think about what you think. But if I can think about your mind, then it seems I should be able to think about my own. Thus, there will be a correlation between those creatures capable of self-reflection and those that resent or stand in the kind of relations vulnerable to the kind of reactions and changes characteristic of moral or interpersonal responsibility. One might then think that the capacity for reflection is required only because it is a correlate of the capacity to think about the reasons of others, a correlate of the ability to think about the quality of others' wills.

This is a tidy solution, but a question remains: Why do those who can think about the reasons of others resent (or have other characteristic reactions) only those who are *also* capable of reflection? Why is the relation symmetric? I can think about my cat's reasons for acting; why is that not enough for me to resent her selfishness and evident disregard for my interests?

Many are again tempted, at this point, to appeal to reflection as affording a kind of control: we do not resent our cats because they are not capable of controlling themselves in the right kind of way. But this is just the thought I have been arguing against all along. To summarize: First, while the kind of control that reflection secures helps to explain, and so is required for, *jurisdictional* responsibility, it is not clear either how would explain or why it should be required for the evaluations and responses characteristic of responsibility—it is not clear why I must be able to *manage* an attitude before it can show kindness or disrespect or license gratitude or resentment. Sometimes people argue that the ability to reflect and change one's attitudes is required for the aptness of reactions like resentment. There are two forms this claim can take. First, resentment (as well as charges of disregard, pettiness, and the like) is a sort of burden, and it can seem unfair to burden someone who lacked an opportunity to avoid the burden. I have argued that this thought is confused. Second, it is sometimes thought that resentment and the like somehow include, within them, a commitment to the claim that the one resented could done better by trying harder, either on

this occasion or in the past. But I have suggested that we need not identify being responsible with being the apt target of attitudes that entail such commitments. Finally, sometimes people argue that moral demands, in particular, do not apply unless their target has the capacity to satisfy them and that, without the ability to reflect, we would not be able to satisfy moral demands. Thus the demands would not apply to us. I have argued that moral demands do not fail to apply simply because an individual is unable to satisfy them.

There is yet another common thought, not yet considered. It is sometimes suggested that the attitudes characteristic of moral or interpersonal relations are a kind of address or communication, and that such an address is *pointless* if directed at those who cannot recognize their significance.[60] But only those who are capable of reflecting on my reasons for resenting, etc., could recognize the significance of my attitudes. My cat cannot. And, again, those who can recognize my reasons are, presumably, also capable of reflecting on their own minds.

While this would secure the desired symmetry, I think this is not quite right—because I do not think that attitudes such as resentment or betrayal are forms of address or communication. In fact, I do not think they are adopted or held for any purpose, communicative or not—any more than a belief is adopted or held for a purpose.[61] If these attitudes were forms of address or communication, they would fail, from pointlessness, if their target is out of the room or far away or

---

[60] Or, at least, they they are enough like address or communication so as to be shown inapt when pointless. Gary Watson introduces the notion of communication in Gary Watson, "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in *Perspectives on Moral Responsibility*, ed. John Martin Fischer and Mark Ravizza (Ithaca: Cornell University Press, 1993). The thought is developed more recently in Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University, 2006); Michael McKenna, *Conversation and Responsibility* (New York: Oxford University Press, 2012). These views emphasize the meaning or significance of action and the correlative meaning or significance of the reactive attitudes. They address the objection I raise in the text by distancing themselves from actual communication. I believe they thus come very close to the position I advocate in the next section, where I consider what is required for one's actions to carry this meaning—what makes respect or disrespect possible at all.

[61] I would grant that they function, in our social life, to convey information—as belief functions, in the life of an individual, to guide behavior. But to say they serve a social function in conveying information is not to say that they are forms of address. Blushing conveys information, and perhaps that is part of its function, but it is not a form of address or communication.

dead. But it is perfectly sensible to continue to resent someone who is absent. Likewise, it seems to me to make good sense to resent(-minus) those who will never learn of your resentment, those who are incapable of recognizing this particular episode of resentment, and individuals incapable of change. So it seems to me we need to look elsewhere to explain why we do not resent our cats.

Rather than look to the abilities of the individual or the pointfulness of attempts at communication, I suggest we consider the expectations and demands whose violation such reactions mark and, especially, the relationships constituted by those expectations. Moral demands, I have suggested, are the demands placed upon us by our need to live peacefully with others. But they are not merely this. One could imagine a "society" where peaceful coexistence is secured simply through the reliable exercise of power—in such a society, individuals regulate their behavior in a peace-preserving way by strategic, rather than moral, reasoning. Violations of the peace-preserving expectations and demands, in such a society, would be seen, not as disrespectful, but simply as imprudent (and perhaps a cause for anger). They would not ground reactions such as resentment or indignation.

What, then, does it take to be capable of showing another respect or disrespect? It seems one must be capable of understanding—and so of either heeding or ignoring—another's standing, status, or rightful claim. Is there a connection between being capable of understanding another's standing, status, or rightful claim and being capable of thinking about the other's mind (or about your own)? The answer depends on what this standing, status, or rightful claim is.[62]

I am drawn to the view that the standing in question is the standing to rightly expect some kind of mutual recognition. But not just any mutual recognition will do—predator and prey might be said to recognize one another, as predator and prey, and, with the right cognitive sophistication, might even expect to be so recognized, yet such recognition would not put them in relations in

[62] If it is simply standing to rightfully declare how other's shall act, then, though you must be able to think about matters of right and authority, you need not be able to think about minds.

which they could be said to respect or disrespect one another or in which resentment or betrayal would make sense. What do we need to add? It seems that, to resent or feel betrayed, you must have expectations that require the other, not just to anticipate your likely behavior (as predator and prey might), but to take, as among her reasons, the fact that you could rightly have expectations of her—that you could rightly ask her to do otherwise. Perhaps, then, the status or standing in question is that of being entitled to such expectations.[63] This would give some content to the idea of recognizing another as "equally real"—we recognize the other as entitled to make claims on us as we are entitled to make claims on the other.[64]

But now we are very close to our target. To recognize another as someone who could rightly have expectations of you, you must be capable of thinking about expectations—but, to think about expectations is to think about another's mind. Creatures who are not able to think about another's mind will also not be capable of either heeding or ignoring another's rightful expectations. They will therefore not be capable of showing either respect or disrespect—not because those creatures are, due to their lack of self-reflection, in some way *out of their own control*, but rather because they do not stand in the right sort of interpersonal relationship with others.[65]

If something in this neighborhood is correct, we would have an account of the correlation between moral responsibility and the capacity for self-reflection, but one which does not appeal to

---

[63] We need some distance between the standing and the expectations—because even people who correctly expect to be *dis*respected have the standing to rightfully expect otherwise (otherwise, it seems, they could not be disrespected).

[64] I am simply taking for granted some notion of entitlement. In the present context, this is licit. The picture of the nature of moral demands is drawn from Scanlon, *What We Owe*. I provide a reading of that work in Pamela Hieronymi, "Of Metaethics and Motivation: The Appeal of Contractualism," in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. Rahul Kumar, Samuel Scheffler, and R. Jay Wallace (Oxford: Oxford University Press, 2011). The picture, as I mean to present it, is one in which the rightful entitlement to expectations of another is not fully grounded in any prior fact of the matter (it is not fully grounded, e.g., in facts about human well-being or the nature of rational wills). Rather, the rightful expectations are rightful because they are those that could be instituted between creatures (with certain specific interests and possibilities for their well being) who are recognizing each other as having rightful expectations. They are expectations possible in a Kingdom of Equals.

[65] One might want to say that pets do come to have expectations—some of which are rightful—and even that they recognize our expectations of them. If so, we stand in a kind of relation continuous, in important ways, with the relations we stand in with other humans. I would take it to be a strength, rather than a deficiency, of the account, if it allows for such continuities.

self-reflection as affording control over the self. Rather, the kind of relationships, expectations, and reactions that constitute us as morally responsible are interpersonal: they are reactions had by and relationships that exist between creatures who can recognize one another's reasons and expectations. The importance of reflection, then, is not in securing control. It is rather a capacity enjoyed by those capable of a certain kind of interpersonal recognition.

7. Conclusion

In conclusion: a common line of thought claims that we are responsible for ourselves and our actions, while less sophisticated creatures are not, because we, and not they, are self-aware. Our self-awareness is thought to provide us with a kind of control over ourselves that they lack.

I have argued that this thought is badly, though subtly, confused. It uses, as its model for the control that grounds our responsibility, the kind of control we exercise over ordinary objects and over our own actions: we represent to ourselves what to do or how to change things, and then we bring about that which we represent. But if there is a question about why or how we are responsible for our actions, it cannot be answered by appeal to a sophisticated, self-directed action. There must be some more fundamental account of how or why we are responsible.

I have suggested a novel but natural replacement: responsible mental activity can be modeled, not as an ordinary action, but as the settling of a question. This requires abandoning the tempting but troublesome thought that responsible activity involves discretion and awareness—which, I have argued, we must abandon in any case.

Finally, I have tried to say something preliminary about why it seems that only creatures capable of self-reflection are regarded as morally responsible. I've suggested that this is because reflection is

required before we can stand in the kind of relationships that constitute us as morally responsible. But this thought, especially, requires more work.[66]

[66] This long-simmering paper has benefited from comments from and conversation with many, including Matthew Boyle, Tyler Burge, Michael Bratman, John Broome, Stephen Darwall, Gerald Dworkin, Ronald Dworkin, David Ebrey, John Fischer, Harry Frankfurt, David Goldman, Mark Greenberg, Steven Gross, Barbara Herman, Paul Hoffman, Robert Hughes, Jenann Ismael, Mark C. Johnson, Sean Kelsey, Aimée Koeplin, John McDowell, Victoria McGeer, Benjamin McMyler, Richard Moran, Thomas Nagel, Philip Pettit, Huw Price, Joseph Raz, T. M. Scanlon, Tamar Schapiro, Samuel Scheffler, Kieran Setiya, Seana Valentine Shiffrin, Sigrún Svavarsdóttir, Tamar Szabo Gendler, Matthew Talbert, Julie Tannenbaum, Sergio Tennenbaum, Gary Watson, Ralph Wedgwood, and Gideon Yaffe. It has greatly benefited from the input of thoughtful audiences at *Agency and Legal/Moral Responsibility*, Antwerp; *Aims and Norms: Action, Belief and Emotion*, University of Southampton; Stanford University; University of Southern California; Northwestern University; *Practical Reasoning: Ancient and Modern Conceptions,* Stanford University; *Colloquium in Legal, Moral, and Political Philosophy,* NYU School of Law; Oxford University; University of Toronto; *Ethics and Moral Psychology, in honor of John McDowell*, University of Sydney; *Values and Valuing*, Reykjavik; *Self-Knowledge and Rational Agency*, Center for the Study of Mind in Nature, Oslo; Texas A&M University; UCLA/USC Graduate Student Conference; University of California, Davis; APA Eastern Division Meeting; University of Washington Graduate Student Conference; *Social and Normative Ethics Workshop*, Stanford University; University of Pittsburgh; Johns Hopkins University; University of California, Riverside; Yale University; *Agency and Action*, Wake Forest University; and *Regulating Attitudes with Reasons*, Dubrovnik. Finally, work on this paper was generously supported by a Frederick Burkhardt Fellowship from the American Council of Learned Societies and by the Center for Advanced Study in the Behavioral Sciences at Stanford University.

BIBLIOGRAPHY

Anscombe, G. E. M. *Intention*. Oxford: Blackwell Publishing Co., 1957.

Boyle, Matthew. "Transparent Self-Knowledge." *Proceedings of the Aristotelian Society* Supplementary Volume 85 (2011): 223–41.

Burge, Tyler. "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96 (1996): 91–116.

Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University, 2006.

Donnellan, Keith S. "Knowing What I Am Doing." *The Journal of Philosophy* 60, no. 14 (Jul. 4 1963): 401–09.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (1969): 828-39.

———. "Freedom of the Will and the Concept of a Person." Chap. 2 In *The Importance of What We Care About*. 11–25. Cambridge: Cambridge University Press, 1988.

———. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (1971): 5-20.

Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55.

———. "Believing at Will." *Canadian Journal of Philosophy, Supplementary Volume* 35 (2009): 149–87.

———. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, no. 1 (March 2006): 45–74.

———. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (2004): 115–48.

———. "Forgiveness, Blame, Reasons..." In *3am: magazine*, edited by Richard Marshall, 2013.

———. "Freedom, Resentment, and the Metaphysics of Morals." (in progress).

———. "The Intuitive Problem of Free Will and Moral Responsibility." (in progress).

———. "Making a Difference." *Social Theory and Practice* 37, no. 1 (2011): 81–94.

———. "Of Metaethics and Motivation: The Appeal of Contractualism." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by Rahul Kumar, Samuel Scheffler and R. Jay Wallace. 101–28. Oxford: Oxford University Press, 2011.

———. "Rational Capacity as a Condition on Blame." *Philosophical Books* 48, no. 2 (April 2007): 109–23.

———. "Reasons for Action." *Proceedings of the Aristotelian Society* 111 (2011): 407–27.

———. "Responsibility for Believing." *Synthese* 161, no. 3 (April 2008): 357–73.

———. "The Will as Reason." *Philosophical Perspectives* 23 (2009): 201–20.

———. "The Wrong Kind of Reason." *The Journal of Philosophy* 102, no. 9 (September 2005): 1–21.

Hobart, R. E. "Free Will as Involving Determination and Inconceivable without It." *Mind* 43 (1934): 1–27.

Kant, Immanuel. *Critique of Practical Reason*. Translated and edited by Mary Gregor. Cambridge Texts in the History of Philosophy. Edited by Karl Ameriks Cambridge: Cambridge University Press, 1997. 1788.

Kavka, Gregory. "The Toxin Puzzle." *Analysis* 43 (1983): 33-36.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

McKenna, Michael. *Conversation and Responsibility*. New York: Oxford University Press, 2012.

Roessler, Johannes, and Naomi Eilan, eds. *Agency and Self-Awareness*. Oxford: Oxford University Press, 2003.

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

Setiya, Kieran. "Explaining Action." *The Philosophical Review* 112, no. 3 (July 2003): 339–93.

Shah, Nishi. "How Truth Governs Belief." *The Philosophical Review* 112 (2003): 447–82.

Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. January (2005): 236–71.

Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75 (1994): 5-24.

Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* xlviii (1962): 1-25.

Velleman, J. David. "On the Aim of Belief." In *The Possibility of Practical Reason*. 244-81. Oxford: Oxford University Press, 2000.

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza. 119-48. Ithaca: Cornell University Press, 1993.

———. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227-48.

Wolf, Susan. *Freedom within Reason*. New York: Oxford University Press, 1990.