

Reflection and Responsibility: Iceland Presentation

Pamela Hieronymi
hieronymi@ucla.edu
June 20, 2010

A common line of thought claims that we are responsible for ourselves and our actions, while less sophisticated creatures are not, because we, but not they, are self-aware. Our self-awareness, it is thought, provides us with a kind of control over ourselves that they lack: we can reflect 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 kind of 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 kind of 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 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 was 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 requires abandoning the tempting but troublesome thought that responsible activity involves discretion and awareness—which, I argue, we must abandon in any case.

THE COMMON LINE OF THOUGHT AND A RESPONSE: IN SKETCH

I begin by roughly sketching the common line of thought together with my response [and, in this format, I will get no further].

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 instances, 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. 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 whatever 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 kind of control is both a ground for and 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.¹

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. When we reflect upon our lives, it seems that each decision we make, and each thing we do, can be adequately explained by conditions in place prior to it. And so, from our reflective vantage, it seems that we do not control our lives: the future, it seems, is already explained by the past, and so, 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 course of history will erode our sense of control over even our own intentional actions.³ And so a sort of threat appears, sparking the free will debate.⁴

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;⁵ 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. ^{6 7 .8}

Each extreme seems unsatisfying. The first requires positing some or another mystery⁹—something like a noumenal self, or a soul working through the pineal gland, whose decisions, though efficacious, are (awkwardly) not (wholly) explicable in terms of the contingent psychology of the empirically given subject.¹⁰ The second avoids the mystery by giving up the claim our responsibility is grounded in and conditioned by some form 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, at least in recent years, 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.¹¹

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 kind of reflective distance and therefore are not responsible for their thoughts or their actions in the way we are.¹²

Again, I believe this reflective strategy is mistaken. It suggests that we control ourselves, most fundamentally, by acting upon ourselves. But, to put it quickly, if there was a question about how or why we are responsible for our intentional actions, we cannot answer that question by appeal to a self-directed intentional action.¹³

The champion of reflection will object that her position is here caricatured. And she might offer a number of different elaborations. First, she might reply that¹⁴ 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.

In reply, I'll grant that there may be such *sui generis* reflective activity and that it may be important for many things.¹⁵ But I would insist 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).¹⁶

To illustrate the lack, I will sketch 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, the appropriate target of certain sorts of reactions on account of it.¹⁷ 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, 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.¹⁸

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 or the the disarray of our apartment. We can be open to assessment on account of the misbehavior of our dog, and, depending on the outcome of that assessment, we may be¹⁹ thought careless, negligent, indulgent, or sentimental; we might be the object of resentment, indignation, outrage, or contempt.

Plausibly, the responsibility we bear for this latter range of things is explained, in part, by appeal to our responsibility for our intentional actions. What responsibility you bear for your dog's behavior derives from the fact that these are things you can affect and so perhaps control through your intentional actions, together with the fact that you are rightly expected to affect and control them in certain ways. 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,²⁰ 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.²¹ Rather, if we can be responsible for things because we can affect and control them through our actions, our responsibility for and control over our actions must be explained in some other way.²²

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.²³ Anscombe noted that, whenever one intentionally ϕ 's, one can rightly be asked, "Why did you ϕ ?" (or "Why are you ϕ -ing?") where this question looks for what she calls a "reason for acting." [I am going to presume we understand, well enough, this question—there is some elaboration on the handout.] Notice that the reasons this question looks for, retrospectively, are just the reasons for which one would, prospectively, settle the question of whether to act. I would suggest, then, that the question is made apt by the assumption that one is ϕ -ing because one has settled the question of whether to ϕ — because it is, in general, true that, if one has settled a question, one can rightly be asked one's reasons for doing so. So I suggest that, whenever one intentionally ϕ 's, one has, in some sense, settled the question of whether to ϕ , and that this settling grounds and explains one's answerability.

This account of answerability will also, I believe, ground and explain our responsibility for intentional actions: In revealing one's positive answer to the question of whether to ϕ , one's intentional actions therein reveal something of one's mind. One's answer to this question will cohere, more or less imperfectly, with other things that one believes and intends, and so reveal²⁴

what one finds worth doing, and, by extension, something of what one thinks true or valuable. If we further know something about the reasons (if any) for one's positive answer, then we will know something more of one's mind.

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 sorts of 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.²⁵

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. 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: we are responsible for our intentional actions because they reveal our answering of a particular question about a particular action in a particular 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.²⁶

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 inward 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.

However, we must note that, if we model our fundamental responsible activity on settling a question, then we cannot enjoy discretion over that which we are fundamentally responsible. Without discretion, I suggest, awareness does not create or enhance control.

Very briefly, on why we do not enjoy discretion: We are fundamentally responsible for a thing, I 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.²⁷ 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 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. ²⁸

Turning, now, to awareness: Notice that, without discretion, the importance of awareness is far less clear. It is unclear why it helps to be *aware of* your take on certain objects, if that the

awareness will not change your take.²⁹ Of course, one reason for wanting to be aware of a thing is to be able to make decisions about how to act with respect to that thing.³⁰ 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. This kind of enhancement of agency is, of course, open to us, in the case of our own mind. It is unclear, though, that lacking an awareness of one's mind in forming an intention or belief leaves one with any less control over whether to intend or believe, or that having it would put one in any more control. So, if the champion of reflection wants to preserve discretion and awareness, she needs to provide an alternative account of responsibility (because the one I sketched here rules out discretion and make awareness unimportant).

The champion of reflection might reply in a different way. She might say that she, too, is interested in a person's take on the world and her place in it, but that reflection is required to reveal the mind—or, at any rate, the *responsible* mind—of the person. Reflection, then, does not so much afford control as it does locate or identify the responsible person. But the question, then, is the one asked, in effect, by Gary Watson, long ago: why should I be identified with my reflecting self and that which that reflecting self endorses?³¹ I hope it clear that we cannot, at *this* point, appeal to the ordinary notion of control—because, again, if it were unproblematic to appeal to it here, we could have appealed to it in the beginning of our inquiry.

A different kind of answer would appeal to reflection as important for securing what I would call *authenticity*. Authenticity is, roughly, the good secured by a liberal arts education and the capacity for critical thinking: someone who can step away from, doubt, and re-evaluate the assumptions which have shaped her is, indeed, more truly her own self. She is also, in a very important sense, liberated; she enjoys a very important form of freedom. However, *this* form of

freedom or identity, I would insist, is not that which is at issue when thinking about responsibility. The inauthentic are nonetheless responsible.³²

Setting aside authenticity, it is striking that responsible creatures seem also to be creatures capable of self-reflection. Here is a very preliminary set of thoughts on why: To be responsible, we've said, is to be the appropriate target of a certain range of reactions, or to be vulnerable to certain changes in an interpersonal relation. But *this* range of reactions, or this kind of relation, 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 for acting, only if I can, so to speak, contemplate your maxims. But if I can think about your reasons for acting, then it seems almost guaranteed that I can also 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 changes characteristic of responsibility.

A question remains: Why do those who can think about the reasons of others resent only those who are *also* capable of reflection? Or, why must the relation be symmetric? Why shouldn't those capable of thinking about the reasons of the cat also resent *her* disregard?

One kind of answer would claim that creature incapable of thinking about their reasons do not act on reasons. I won't pursue that. Another, tempting, thought appeals, at *this* point, to reflection as affording a kind of control: we do not resent cats because they are not capable of controlling themselves in the right kind of way. I would resist this in two ways. First, I would note that the kind of control that reflection secures explains only jurisdictional responsibility. It does not explain answerability. And it is unclear why jurisdictional responsibility should be a precondition for answerability: it is unclear why I must be able to manage an attitude before I am

answerable for it. Second, I would question the motivation for this view. The underlying thought seems to be one of fairness. But I have elsewhere suggested this is mistaken.

I would pursue a very different line. Recall that the capacity for self-reflection came with the capacity to think about the reasons of others. Better, then, to ask, “Why will creatures capable of thinking about the reasons of others resent only those who are *also* capable of thinking about the reasons of others?” And here I think we are closer to an answer: only a creature capable of recognizing the reasons of others is capable of recognizing my reasons for resenting, and so capable of recognizing the protest my resentment marks. My resentment, it seems, makes sense only if directed to those capable of recognizing why I resent.

But why should this be? Why should resentment makes sense only if directed at those capable of recognizing its basis? Some would answer that resentment is a kind of address, and so pointless if directed at those who cannot recognize its basis. But I think this is not quite right—because I don’t think that resentment has or needs a point, any more than a belief does. I think it makes sense, e.g., to resent those incapable of recognizing this *particular* episode of resentment, or those incapable of change. Rather, I think that resentment marks the violation of expectations that are, roughly speaking, instituted between creatures capable of recognizing the quality of each other’s will, and so capable of demanding mutual respect.³³

If this is right, we have an account of the *correlation* between responsibility and the capacity for self-reflection, but one which does not appeal to self-reflection as affording control over the self. Rather, the kind of relationships, expectations, and reactions that constitute us as responsible are interpersonal: they are reactions had by creatures who can recognize one

another's reasons. 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.

[The free will debate will now look different.]