

Looking for Control

Peter Eichman

December 2005

1 The Problem

It seems like common sense to say that we are morally responsible only for those things which are under our control. But there are also many cases where we feel justified in assigning blame (or praise), yet the agent who we are blaming was not in control of the action that we are blaming them for. At first this may not be readily apparent, but it soon becomes so once we consider all of the ways in which factors beyond our control influence our actions and the results thereof. This is what is at the heart of the problem of moral luck:

Moral luck occurs when an agent can be correctly treated as an object of moral judgment, despite the fact that a significant aspect of what he is assessed for depends on factors beyond his control. (Nelkin, Spring 2004)

A possible source of the problem is that there is a fundamental conflict between a theory of actions where actions are reducible to mere events, and the conception of agency (Nagel, 1979). If an action is reducible to an event, and if every event is caused by some other event, then there seems to be no more room for agency. Taken to this extreme, it seems we would no longer have agents, and our notion of moral responsibility really starts to crumble. Since it seems like moral responsibility requires control, and control (to a

first approximation) implies agency, if the concept of agency disappears under this sort of theory of action, then so does the concept of moral responsibility.

In this paper I will examine the question of how we determine responsibility. I will assume (primarily on intuitive grounds) that one of the necessary conditions of moral responsibility is control. Given that, I examine how that notion of control might be explained in terms of more basic mental objects (desires, intentions, and so forth). Along the way, I will be building up a definition of moral responsibility that attempts to capture our common sense notion of when we are justified in blaming (or praising) people. But ultimately, I believe it comes down to the notion of control: I will argue that we may not have the control we think we have, or if we do, it is not in the place we think it is.

Kant is often cited as the ultimate source for the thesis that control is necessary for moral responsibility. In the *Grounding for a Metaphysics of Morals*, he asserts that:

A good will is good not because of what it effects or accomplishes, nor because of its fitness to attain some proposed end; it is good only through its willing, i.e., it is good in itself. When it is considered in itself, then it is to be esteemed very much higher than anything which it might ever bring about merely in order to favor some inclination, or even the sum total of all inclinations. Even if, by some especially unfortunate fate or by the niggardly provision of nature, this will should be wholly lacking in the power to accomplish its purpose; if with the greatest effort it should yet achieve nothing, and only the good will should remain (not, to be sure, as a mere wish but as the summoning of all the means in our power), yet it would, like a jewel, still shine by its own light as something which has its full value in itself. Its usefulness or fruitlessness can neither augment nor diminish this value. (Kant, 1785/1993, pp.7-8/GMS 394)

By coupling that with Kant's stance that the proper object of moral assessment is the will, and not any particular action by an agent, we arrive at a view that moral responsibility should not be subject to luck¹. This leads us to the Control Principle:

We are morally assessable only to the extent that what we are assessed for depends on factors under our control. (Nelkin, Spring 2004)

There is a great intuitive appeal to it; again it just seems like common sense. Conversely, if I am not in control of my actions, I should not be blamed, even if they bring about a negative outcome.

Nagel and others agree that control is required for moral responsibility. However, Nagel for one argues that once we start to seriously consider the influence of luck on our lives, the sphere of things under our control shrinks "to an extensionless point" (Nagel, 1979, p.35). This then is the crux of the moral luck problem. Either we are morally responsible only for those things which are under our control, or else we are morally responsible things that are beyond our control. The first option leaves us with nothing for which we are morally responsible, while the second seems to fly in the face of the common sense notion that control is central to moral responsibility.

Though this begins with a Kantian principle of control being necessary for moral responsibility, given that the will is the ultimate object of moral assessment, we must ask how are we to know whether a will is good or bad, except by the actions that are effected by the possessor of that will? Actions are our only evidence of will, so we naturally use these

¹For the duration of this paper, I will be using the term 'luck' simply as a rough synonym for "not under (the agent's) control." I emphatically do not intend it in any pejorative way, or in a way that implies the sense of "unlikely" or "against the odds" that is more more common to everyday speech. I only point this out because, although the majority of the moral luck literature I have looked at sticks to this usage, the everyday sense of the word often creeps in and can at times confuse the issue.

to judge the will of the actor. Protestations made by the subject of such judgments to the effect that what they did was out of character, and that they really do have a good will² only remain effective as long as the bad actions by that person remain an aberration. As soon as we sense a pattern of bad actions, we come to believe that the person in question must really have a bad will. On the flip side, we do sometimes judge the moral worth of actions independently of their agents. We say things like, “That was a bad thing to do, but I don’t consider you a bad person because of it.” But again, once we see a pattern, I think the object our judgment shifts from the action to the actor.

All this is to say is that though the current project may have roots in the notion of the will as the seat of moral responsibility, for practical purposes we must look to the actions of agents in order to determine the nature of their will. And it is in these execution of these actions that moral luck comes into play. Let us quickly review the four forms of moral luck that Nagel lays out. They are consequential, circumstantial, constitutive, and causal luck. We are subject to consequential luck because factors beyond our control often contribute to the success or failure of our actions to arrive at our desired ends. In other words, the world may or may not cooperate with our intentions. We are subject to circumstantial luck because the particular situations we find ourselves in (and hence the moral choices we are able to make) are not under our control. We are subject to constitutive luck because the make up of our character and personality, and thus the sorts of whims and desires we tend to have, is not under our control. Much of it is a matter of genetics (which is certainly beyond our control) and upbringing (which is also, for all practical purposes beyond our control, at least as young children). Finally, we have causal luck. This is the hardest of the four to get a conceptual handle on. It has to do with the scope of what one is morally responsible for. Even granting that we have control over our “stripped-down acts of will,”

²Or, perversely, a bad or evil will. Though I doubt this happens much if at all in real life, I have seen it used (often for comic effect) in fiction.

because of the other three forms of moral luck, we are not in control of which acts we can end up being responsible for.

One important point to notice is the distinction between moral responsibility and (mere) causal responsibility, and not to confuse them (Enoch and Marmor, Forthcoming). Causal responsibility does not require the agent to be in control of their actions. Causal responsibility is a much broader notion; there are certainly things that I am causally responsible for that I am not morally responsible for. Nor is it the case that moral responsibility is simply a subset of causal responsibility. There are cases where we may hold a person morally responsible for an event even if they are not causally responsible for that event. The paradigm case for this would be cases where commanding officers in a military are held responsible for the actions of their subordinates, even when there is no causal link between the superior and the specific actions of the subordinates.

2 Defining Control

What we need now is a more precise spelling out of the notion of control. Zimmerman has one account, where he splits control into two types, restricted and unrestricted:

Restricted Control An agent is in “restricted control with respect to an event just in case one can bring about its occurrence and can also prevent its occurrence” (Zimmerman, 1993, p.219)

Unrestricted Control An agent is in “unrestricted or complete control with respect to some event just in case one [is or was in] restricted control with respect both to it and to all those events on which its occurrence is contingent” (Zimmerman, 1993, p.219)

Note that it seems that no one could ever be in unrestricted control of any of their actions. For any action that I perform, that action is contingent on the event of my being born. But I am not in control (restricted or otherwise) of my being born³. Even Zimmerman himself admits that it is true that “no event is such that anyone is ever in unrestricted control of it” (Zimmerman, 1993, p.220-221).

That leaves restricted control, where the agent can either bring about or prevent an event from happening. It would be nice to have a more precise definition of what it means to be able to bring about or prevent an event. Let us begin by seeing how intention might be of aid in defining control. If I am in control of something, then it seems plausible to say that I either intend to do that thing, or I intend to not do that thing. Also, whether I take action or refrain from doing so will be connected to my intention. That is to say that S is in control of ϕ just in case S does ϕ if and only if S intends to ϕ .

We have to be careful here in order to avoid a certain regress. I have just defined control in terms of intention. But it is perfectly sensible to ask how are the formations of one’s intentions controlled? It surely feels like we have control over what intentions we have. But if control is simply reduced to intention, then we arrive at a regress of intention. In response, I will claim that the sort of control we exercise when acting is different from the sort of control we exercise when forming intentions. The former is reducible to intentions, while the latter is something else entirely⁴.

Note too that the above merely talks about ϕ , some action that the agent S performs, while Zimmerman’s definition of control (and ultimately, any sensible definition of moral responsibility) talks about a broader class of *events* that S is responsible for. This I take

³And to people who would argue with that, the contingent event can be pushed back even farther: my being conceived, my mother being born, the planet Earth coming into existence, the Big Bang, and so on, until we find an event that we can agree that I had no control over

⁴See Section 3 for my thoughts on what that something might be

to include events that are not actions performed by S . Instead, some of these events would be results or consequences of actions performed by S . To explicitly connect these, then, I would say that S is morally responsible for some event e just in case S is in control of ϕ and ϕ causes e .

So far this gives us a sort of restricted causal responsibility, where I am responsible for any of the results or consequences of the actions which are under my control. This is no longer as broad as pure causal control, since we have restricted it to those actions which are under my control. But in a sense, this is still too broad. It still admits what I call “chaos theory cases”⁵; these are cases where some (intentional and fully controlled) action of mine causes a disturbance in the environment that is a necessary cause of some distant (and usually portrayed as catastrophic) event. For example, the motion of air that I cause by walking from my car to campus may make a causal contribution to a weather system that eventually forms into a hurricane that damages buildings in a town on the gulf coast of Florida. That this happens is a brute fact about the world, explicable by laws of nature. However, it seems ludicrous to say that I am morally responsible for the hurricane, the damage, even any loss of life that may accompany it (see also Enoch and Marmor (Forthcoming)). This cannot be just an issue of S not being in control of ϕ , because I was certainly in control of my initial action (walking to campus). It is just that the consequences of that action have somehow spiraled out of my knowledge and control.

In addition to control, moral responsibility also seems to require a certain amount of knowledge on the part of the agent. As a first approximation we would like to say, the agent S must know that their action ϕ will cause the event e in order to be morally responsible for e . However, since it is not practically possible to know *that* ϕ will in fact cause e , we weaken the requirement to S knowing that it is likely that ϕ will cause e .

⁵When a butterfly flaps its wings...

In point of fact, we may even go further and make that requirement into a normative requirement: *S should know* that it is likely that ϕ will cause *e*. This prevents the idiot from being the most morally good person around, not in virtue of their actions, but in virtue of the fact that they do not know what the likely results of their actions are. This is also an encoding of the maxim (often applied in the context of law) that “ignorance is no excuse.”

Note that I do not think that the absurdity of the “chaos theory case” has to do with our inability to know (due to the inherent chaotic nature of such things as weather systems) all of the effects that our actions will cause. Instead I think there is a different practical problem. Even if we were able to map out all the ways in which we are causally related to the world at large, I still think we would have a problem equating causal responsibility with moral responsibility. The problem is this: how do we avoid any untoward effects of our actions. The simple solution would seem to be, avoid those actions which we case see have untoward effects. In fact this is how we normally operate, it is just that our scope of foreseeable consequences is usually pretty limited. I have the feeling that if we were able to have a perfect causal understanding of the world, and we were to then equate moral and causal responsibility, we would become paralyzed and unable to act⁶, because we would see the causal connection between all of the (possibly distant) negative effects and any of our actions.

Another problem I see arising with our current definition of moral responsibility involves the case of the accidental poisoner:

- (1) *S*, not knowing that the sugar in the sugar bowl has been replaced with arsenic, puts a spoonful of it into his friend’s coffee. The friend dies from the poison.

⁶Of course, this is assuming we wish to continue doing the morally right thing. If not, all bets are off

In this case, the intuition is that S should not be held morally responsible for poisoning his friend. Also, I think we find it intuitively plausible to say that S is not in control of putting arsenic in the coffee. However, according to most theories of action, the act of putting the white powder into the coffee *just is* the act of putting poison in the coffee. Since S intended to put the white powder in the coffee, he simultaneously intended to put the arsenic in the coffee, so he was in control of putting the arsenic in the coffee. By requiring that the description of the act ϕ must be one that S would assent to at the time that he is doing ϕ , we can avoid this, and better capture our intuitions in this case.

I think this case highlights something central to the notion of moral responsibility, as opposed to causal responsibility. Moral responsibility is not a fact about the world in the way that causal responsibility is. Even if we were not around to make judgments about things, certain events would still be causally responsible for other events. However, without human beings, or more broadly any other sort of moral agents, there would not be any facts about the moral responsibility for actions. In this way, moral responsibility is more akin to knowledge than causation, with all of the problems of evidence that that brings with it (Nagel, 1979). There is linguistic evidence for this. Notice that we can have either a person or an event the subject of the sentence “ S is causally responsible for e ,” but only a person⁷ can be the subject of the sentence “ S is morally responsible for e .”

At this point, we have a definition of moral responsibility whereby an agent S is morally responsible for an event e just in case:

1. S is in control of some action of theirs ϕ
2. ϕ must be stated in terms that S would assent to at the time of ϕ -ing
3. S knows that it is likely that ϕ will cause e at the time of their ϕ -ing

⁷And not only that, but they must be capable of being considered a moral agent, whatever that may entail.

4. $S \phi$ -s
5. ϕ causes e .

3 On the Cognitive Nature of the Will

Even if we grant that luck may play a factor in the circumstances we find ourselves in, our natures and dispositions, and in the actual results of any action we perform, the Kantian in us may continue to hold up the very acts of will by which we act as the true object of moral assessment. After all, these are supposed to be the things that are truly under our control.

There is one form of luck here, though I am not sure how problematic it is in comparison to the others. Even though we may have a completely free will (i.e., we are able to freely decide between any possible actions), it is still a matter of luck that we have this very free will. It certainly was not under our control that we received or attained this faculty.

I am also drawn to the question of what exactly the will is. This is certainly not an easy question, and I am sure that I will not come up with a definitive answer in this short space. However, I would like to sketch a sort cognitive of architecture of the mind and especially of the will, that I believe has implications for questions regarding moral responsibility and the effect of exterior or deterministic factors on us.

Within our minds is a collection of beliefs, desires, and other mental objects. At some point or another (perhaps provoked by some external stimulus, perhaps not), one of these desires gets singled out as something that we *really* want to do. At this point, we reflect on that desire. Loosely speaking, we are evaluating it for a certain sort of fitness, given our other beliefs, desires, intentions, and so forth. If this evaluation comes out positive, the desire becomes an intention; if not, then it remains a desire.

None of this need happen consciously. Indeed, for certain desires and intentions that I will get to shortly, it may be fairly important that this process goes on non-consciously. Certainly, desires of a high level of abstraction are probably processed and refined consciously, but it seems equally likely that there would be desires for which this process of evaluation has become automatic and hence (usually) non-conscious.

Once this initial intention is formed from the desire, it becomes part of the general pool of available mental objects. It will activate related beliefs, desires, and other mental content. Of these, some of the desires will then present themselves to the same evaluation process as the initial desire. More intentions may then be formed, which feed back into the mental object pool, and so on. In essence, it is a sort of positive feedback loop by which the initial desire/intention accretes a whole related set of beliefs, sub-intentions, and so forth.

I speculate that this process would progress in a sort of top-down fashion, beginning with a (usually conscious) desire that is expressed in fairly high level of abstraction (abstraction at least with respect to, say, the actual motor movements required to carry out the fulfillment of said desire). This desire may be something amoral, like “I want to go to the store,” or its content may carry moral weight, as in “I want to kill Phil.” Let us take the second example, and examine a bit further how the path from desire to intention to action might occur, noting where what we call the will is situated within this process.

Say I have the desire D_0 , which can be expressed as “I want to kill Phil.” I evaluate this desire and find that I believe I am able to kill Phil, and that I have no (or at least not enough) qualms about carrying through with that act, and so forth. Now I have the intention I_0 , roughly expressed as “I intend to kill Phil.” This intention then interacts with the other contents of my mind, and I embark on the task of planning my action. Perhaps one of the beliefs that occurs to me is the best way to kill Phil will be to shoot

him. From that arises a new desire, D_I , expressed as “I need/want a gun,” which I can run through the evaluation process and arrive at intention I_I , “I intend to get a gun.”

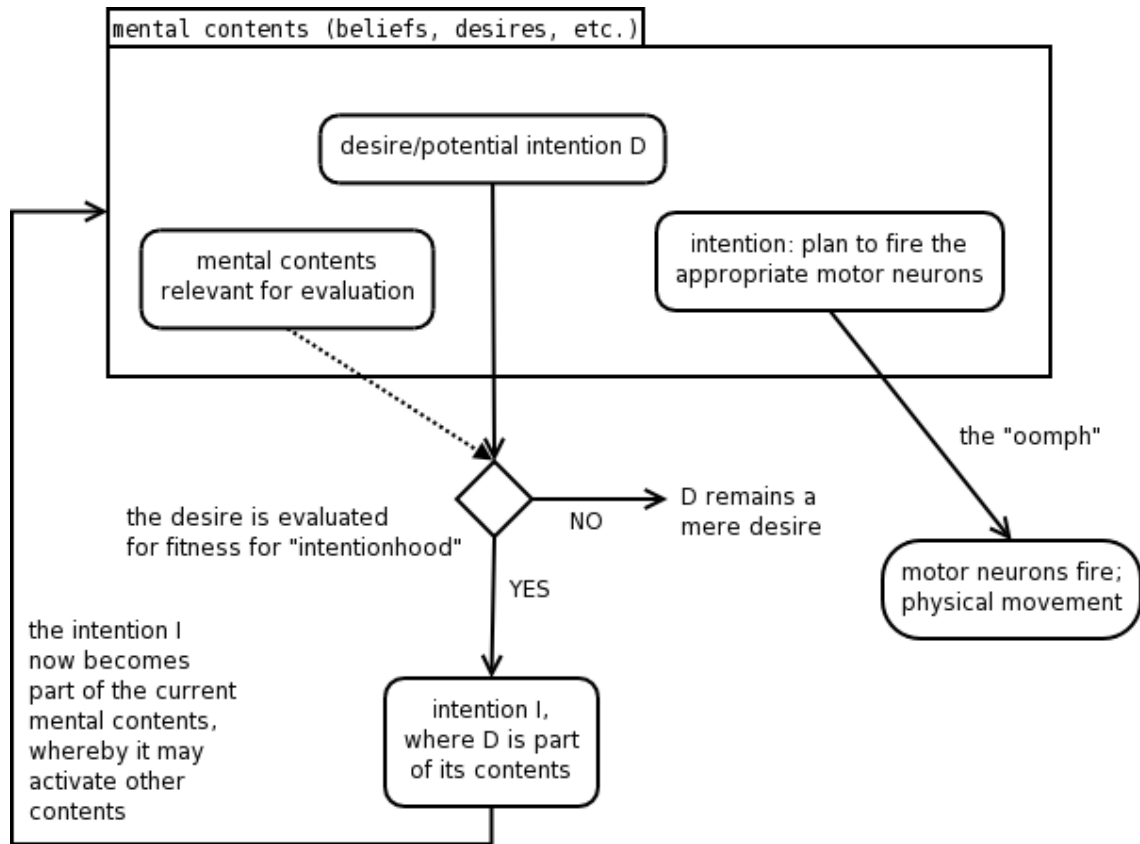
Eventually I will arrive at some intention I_n which I cannot consciously decompose into stages the way I have those that come before. That will probably be something like “I intend to stand up (in order to walk to the pawn shop (in order to buy a gun, etc.)).” Here is where the non-conscious version of this process I alluded to earlier would come into play.

It certainly feels like there is a gap between the intention I_n and the action that I perform. This is where the will is often thought to reside, where the will is taken to be that special (and somewhat mysterious) “oomph” that connects the mental to the physical. I submit that it may not be as mysterious as it seems, and that some of the mysteriousness may be explainable by appeal to ideas of non-conscious processes of the mind.

Again, I hold that the evaluation feedback loop described above continues apace, though the mental contents here may have stopped looking quite like what we generally think of desires and intentions looking like. At this level, the feedback loop is more focused on planning bodily motion, by breaking down desired movements through space into smaller motions, muscle movements, and ultimately nerve firings to control those muscles. This is where I contend that the “oomph” of the will resides. We are conscious of our intention to stand up. This is translated automatically by our mind/brain into the individual nerve firings that compel our legs to move in such a way, for our arms to counterbalance, and so on, and *voilà*, we stand up. To us, it feels like the thought “stand up” immediately impelled the action of standing up.

But now, it seems like the will is no longer something we actually control, especially if we mean by the will that which connects the mental and the physical. On the account I have sketched, that mysterious link has been reduced to the firing of some motor neurons, prompted by a certain sort of non-conscious (and at least partially non-consciously formed)

Figure 1: A possible cognitive architecture of the will



intention. It sounds like nothing could be farther from the notion of control than that.

The ready objections to this view that arise from this concern are really two-fold. First, what of that feeling we have that we are in control? By that feeling, I mean the particular phenomenology that accompanies the actions we choose to do. To this objection, I would respond that there is already a precedent within psychology and philosophy of mind for phenomenological experiences, while opaque to introspection, turn out to have psychophysical descriptions not dissimilar in character to the above characterization of the will. The second and ultimately the more important objection, though, is that by this description I do away with free will, moral agency, and moral responsibility in one broad stroke. Moral

luck wins, in a deep (and somewhat disturbing) sense.

It is not quite this bleak, though. Even within the feedback model, there is yet a little room for some autonomous control. I have not described (nor do I intend to attempt to here) just what sort of thing the evaluation procedure for potential intentions is. That could conceivably be the true seat of control and free will. If this is the case, then when we speak of willing, we mean that mental-physical link instantiated as the firing of motor neurons by the appropriate sort of intention, while when we speak of the will, we mean the free evaluation process that can act as a throttle on that link. And we remain in control of that throttle.

4 Conclusions

In the end, however, I am in agreement with Nagel that the problem of moral luck is one which “has no solution, because something in the idea of agency is incompatible with actions being events, or people being things” (Nagel, 1979, p.37). In addition, if the cognitive structure of that which we call the will is anything like I have described above, then there may not even be an “extensionless point” of moral responsibility. Instead, moral responsibility as we traditionally conceive of it would seem to totally disappear under a strict interpretation of that description of the will as a deterministic system.

Still we remain intuitively convinced that we are the agents in control, and therefore morally responsible for our actions. Yet how can this be rational, if there is moral luck to the extent that Nagel argues? How can it be rational to behave as if we have control of our actions when we are faced with evidence that in fact, those actions and the events they caused are controlled in a large part (or even entirely) not by our will, but by factors external and beyond our control?

I think that those questions are essentially misguided in a deterministic scenario. This

does not mean that there are no interesting or useful questions to be asked. Remember that no matter what the ultimate outcome of the question of determinism is, what remains in us is the experience of agency, the experience as of control. And here is where the analogy to epistemology rather than metaphysics comes into play.

If control over action is described as a phenomenological experience, then we can still (on the surface) ask, was S in control of their action. But notice that this now becomes not a metaphysical question, but a sort of quasi-epistemic or phenomenological question. It is akin to asking if someone is in pain; we are wondering if they are having or have had a certain sort of experience.

References

- Enoch, David, and Andrei Marmor. Forthcoming. The case against moral luck .
- Kant, Immanuel. 1785/1993. *Grounding for the metaphysics of morals*. Indianapolis: Hackett. Translated by James W. Ellington.
- Nagel, Thomas. 1979. Moral luck. In *Mortal questions*, 24–38. Cambridge: Cambridge University Press.
- Nelkin, Dana K. Spring 2004. Moral luck. In *The stanford encyclopedia of philosophy*, ed. Edward N. Zalta. URL <http://plato.stanford.edu/archives/spr2004/entries/moral-luck/>.
- Zimmerman, Michael J. 1993. Luck and moral responsibility. In *Moral luck*, ed. Daniel Statman, 217–234. Albany, NY: State University of New York Press.