

Thomas Schmidt

How to Understand the Problem of Moral Luck

1. Acts of negligence which do not lead to bad results do not appear to be morally on a par with otherwise identical instances of negligence which cause significant harm to others. The same is true with regard to similar cases in which we at least seem to make our moral assessment of agents dependent on factors which are not, or not entirely, up to them. A much-discussed example is the case of reckless driving. If a reckless driver, in addition to and as a consequence of his driving recklessly, hits a child, this can hardly be denied to be a morally significant matter. Plausibly, however, whether a child crosses the road right in front of the driver is something beyond his control.

Bernard Williams, in his paper “Moral Luck” (Williams 1976), and Thomas Nagel, in his reply to Williams which was published under the same title (Nagel 1976), have both famously argued that these and similar moral phenomena are philosophically significant. Following Nagel, many see them as giving rise to what has come to be called the *problem of moral luck*. According to the by now canonical way of putting this problem, there is a tension between the *control principle*, i.e. the claim that

people cannot be morally assessed for [...] what is due to factors beyond their control (Nagel 1976, 25)

and our moral practice if, as it seems plausible to many, upholding this very practice commits us to believing that there are instances of *moral luck*, i.e. cases in which that for which agents are to be morally assessed at least partly depends on factors which are not, or at least not fully, within their control.¹

¹ Dana Nelkin is among the many who conceive of the claim that there is moral luck, *cum grano salis*, in this way (Nelkin 2008, § 1). Nagel, however, uses the term *moral luck* in order to refer to situations in which “a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgment” (Nagel 1976, 26). Nagel, that is, does not require that the relevant moral assessment is *appropriate*. Under his understanding, moral luck is involved whenever we, as a matter of fact, do make differences in our moral assessments of the sort in question, regardless of whether making these differences is justifiable. Given this understanding, there is no contradiction between the control principle and the claim that there is moral luck. Rather, the control principle and the claim that there is moral luck together imply that our moral practice is, as far as the control principle is concerned, not as it ought to be.

A number of writers believe that this is a deep and unsettling problem of moral philosophy which requires specific philosophical attention. Not all agree, however. A line of thought which has already been pursued by Nagel seems to imply that the problem of moral luck is not a genuine problem in its own right, but that it instead boils down to issues which are addressed in the free will debate.

The main objective of this paper is to show that this claim is false. Moreover, it will turn out that if one does not appreciate that the problem of moral luck is to be distinguished from issues about free will, then one arrives at a distorted picture of the issues which have to be addressed, of the theoretical resources which need to be brought into play, and of the theoretical options which are plausible when it comes to tackling the problem of moral luck.

2. Nagel suggests that if the claim that moral blameworthiness does not depend on results is defended with reference to the fact that results are, at least partly, influenced by luck, then it is difficult to see why the very same thought would not also apply to all other aspects of our agency which also depend on factors which are subject to luck.² It would then follow, however, that “nothing or almost nothing about what a person does seems to be under his control” (Nagel 1976, 26).

According to Nagel, if we accept the idea that blameworthiness only depends on what is free from luck in the sense of not being influenced by external factors, then we are forced to accept that the “area of genuine agency, and therefore of legitimate moral judgement, seems to shrink [...] to an extensionless point” (Nagel 1976, 35). The core of the reasoning leading to this result is as follows:

- (1) Blameworthiness does not depend on results (since results are not within the scope of control).
- (2) If blameworthiness does not depend on results, then we are not blameworthy for anything (if results are not within the scope of control, then nothing is, and blameworthiness depends on control).
- (3) Therefore there is nothing for which we are blameworthy.

² Nagel claims, in other words, that *resultant luck* (i.e. “luck in the way one’s actions and projects turn out”) is, as far as its relevance for blameworthiness is concerned, not different from *circumstantial luck* (or, as Nagel calls it, “luck in one’s circumstances”) which occurs when the circumstances of one’s actions were such that one acts in non- (or less) objectionable ways, nor from *constitutive luck* (i.e. luck with regard to “the kind of person you are”), nor from *causal luck* (“luck in how one is determined by antecedent circumstances”, all quotes in Nagel 1976, 28).

Nagel does not accept the conclusion of this *modus ponens* but believes the situation to be paradoxical (Nagel 1976, 27). As he sees it, the view that this is the case

is not a *mistake*, ethical or logical, but a perception of one of the ways in which the intuitively acceptable conditions of moral judgment threaten to undermine it all. (Nagel 1976, 27)

3. Michael Moore, in his 1994 paper “The Independent Moral Significance of Wrongdoing”, turns the *modus ponens* just presented into a *modus tollens*, thus arriving at a *reductio* of the claim that results do not affect blameworthiness. Suppose this claim were true for the reason that results are beyond our control due to the possibility of intervening causal chains. Intentional bodily movements, rather than events further down the causal road, then appear to be better suited as candidates for what is under our control. Yet with regard to the relation between bodily movements and their causal antecedents, the situation is, says Moore, in relevant aspects analogous to the relation between consequences and bodily movements:

Just as a number of events could intervene between the movement of the trigger finger and the death of the victim that would rob the bullet of its normal capacity to kill, so a number of events could intervene between each of these mental states [sc. the mental states causing the movement of the trigger finger] and their normal effects so as to rob them of their normal causal power. (Moore 1994, 272)

Moving backwards in the chain of the mental states causing our actions ultimately leaves us, so the argument goes, with nothing at all which is in our control and, therefore, with nothing for which we are blameworthy. This, says Moore, is an unacceptable conclusion, and he therefore rejects the initial assumption of the argument. Results, then, do matter to blameworthiness. Rather than inferring (3) from (1) and (2), Moore infers not-(1) from (2) and not-(3).

4. Whether one has no alternative to going for either the *modus ponens* from (1) and (2) to (3) or for the *modus tollens* from (2) and not-(3) to not-(1) depends on whether premise (2) is true. Nagel and Moore back up this premise by a reflection on the notion of control which seems to establish that, with regard to control, there are two and only two options:

Either, the fact that an aspect of an agent’s action is at least partly due to causal factors which are not up to him implies that the aspect in question is beyond the agent’s control. Then nothing is under an agent’s control, since for any action there are aspects causally relevant to it which are not up to him. When conjoined with the control principle, this implies that there is nothing at all

for which agents could be morally responsible and blameworthy and, in particular, that blameworthiness does not depend on results (in other words, (1) and (3) come out as true). In our moral practice, however, we nevertheless do hold one another, at least by and large, responsible for at least some aspects of what we are doing. The ‘problem of moral luck’, accordingly, would not be a tension between *some* of our ascriptions of moral responsibility and blameworthiness on the one hand and the control principle on the other, but between this principle and our *whole* practice of ascribing responsibility and blameworthiness.

Or, it is possible (and, indeed, often the case) that we are in control of both our actions and their results (in other words, (1) and (3) are false), even though both are, at least partly, caused by factors external to us. Then taking seriously our moral practice does not seem to commit us to the claim that there is moral luck, since the scope of control is wide enough such that there is no reason even to suspect that results are, in principle, *beyond* what we can control. Consequently there is no tension worth being called the ‘problem of moral luck’ at all.

If there were no other options than the ones just described, then there is *either* a tension between the control principle and *all* of our ascriptions of responsibility and blameworthiness, *or* there is no deep conflict between the control principle and *any* of our ordinary ascriptions of responsibility and blameworthiness (at least to the extent to which they are based on the view that results are among the sorts of things which can influence blameworthiness). If these were the only options, then there would be no *specific* problem of moral luck, since either *all*, or *none* of our ordinary practice would be dubious, as far as the control principle is concerned.

Observe that these options are exactly the standard ones in the traditional debate about free will and responsibility, if determinism is assumed to be true: if incompatibilists are right in claiming that determinism rules out control, then the control principle implies that none of our ascriptions of responsibility and blameworthiness are justified. If compatibilists are right, then there is no conflict between determinism and control, and consequently no fundamental problem with any of our standard ways of ascribing responsibility and blameworthiness, as far as the control principle is concerned. This is, indeed, the way Moore views matters. Accordingly, he writes, somewhat harshly, that there

is nothing new going on in the badly labelled ‘moral luck’ debate not already going on (for centuries) in the free-will debate. (Moore 2009, 25)³

³ Moore’s view that there is no independent problem of moral luck is also reflected by his

5. If what is commonly referred to as the ‘problem of moral luck’ were to collapse into issues concerning compatibilist and incompatibilist conceptions of free will and responsibility in general, then this would have significant consequences for assessing the standard options for dealing with it which have been proposed in the literature.

If the control principle is upheld, then – on pain of inconsistency – one needs to deny that there is moral luck. There are three ways of harmonizing this with the fact that our moral practice seems to suggest otherwise:

Reinterpretation. The phenomena which might be taken to indicate that there is moral luck can, and ought to be, interpreted such that they are consistent with the denial of moral luck. One might, for instance, be able to explain and justify differences in reactions by pointing to epistemic asymmetries. When something bad has happened as a consequence of some act, this might be taken to be an *indicator* for (a greater degree of) blameworthiness rather than as something which *in itself* is able to influence the degree of blameworthiness. Also, part of the relevant phenomena could be interpreted, to use a term proposed by David Enoch and Andrei Marmor, as “blame-related reactions” (Enoch/Marmor 2007, 412), the justification of which might consist not only in considerations about blameworthiness but also involve, for instance, reference to pragmatic reasons (see Enoch/Marmor 2007, 413 ff.).

Revision. Given certain aspects of our moral practice, we are indeed committed to the view that there is moral luck. Since, in these respects, our practice is not as the control principle says it ought to be, our practice is in need of revision.

Wide control. As with *revision*, the relevant differences in reactions are indeed grounded in the thought that results can *in themselves* be relevant for the degree to which an agent is blameworthy. This, however, does not conflict with the control principle, since, understood correctly, no relevant aspect of human action is *ex ante* excluded from the scope of control.

Going for the *modus tollens* from (2) and not-(3) to not-(1) amounts to adopting the *wide control* strategy. If this were the right view about control, then there

calling the two above distinguished views about control ‘incompatibilist’ and ‘compatibilist’ (Moore 2009, 24–25) (see also § 7 below).

would be no tension between the control principle and our practice of assuming that blameworthiness depends on results, since these are not in principle excluded from the scope of control.

If one believes that, according to an appropriate understanding of control, we do not control anything at all, then one is committed to the *modus ponens* version of the reasoning introduced above, i.e. to inferring (3) from (1) and (2). Then one can, in general, either opt for *reinterpretation* or for *revision*. Reinterpreting our differing reactions in the way indicated above might indeed seem like a promising approach with respect to moral distinctions we make given differences in results. It is, however, difficult to see how *all* of our relevant reactions could be interpreted as being grounded in something other than judgements about responsibility and blameworthiness. This effectively leaves the proponent of the *modus ponens* version of the above reasoning with the *revision* strategy. Yet revising our whole practice with regard to ascriptions of responsibility and blameworthiness such that it conforms to the control principle would – under the assumption that nothing at all is under our control – amount to giving up that practice entirely. Even if this were possible (which is highly doubtful), it would be a move more drastic than most people would be able to accept (see, in particular, Strawson 1962).

These considerations can be seen as corroborating Moore's view that the above *modus tollens*, rather than the *modus ponens*, is the most promising option to go for if premise (2) is accepted. Given that results are not in principle excluded from the scope of what agents can control, and given that the only alternative is that we can control nothing at all, the former is definitively the more attractive view.

6. To summarize: If one thinks about control in the way Nagel and Moore do, then one is committed to accept premise (2). As a consequence, one is forced either to accept (1) and (3), or to reject both. In both cases, there is no moral luck, and the most pressing issues are those discussed in the debate about free will.

Whether there is an alternative understanding of the problem of moral luck to the one forced on us by Nagel's and Moore's reasoning crucially depends on whether premise (2) can successfully be criticized. In what follows, it will turn out that this can indeed be done: even if agents do not control results, it does not follow that they do not control anything at all.

Suppose that R is the relation which needs to hold between an agent A and some X in order for A to have control over X. Without presupposing any specification of R, a generic notion of control can be defined as follows:

An agent A has *control* over some X if, and only if, A stands in relation R to X.

With this notion of control in place, a notion of partial control can be defined as follows:

An agent A has *partial control* over some X if, and only if, A stands in relation R to at least one (or sufficiently many) of the factors which are causally relevant for the occurrence of X.⁴

The scope of control depends of course on the control-constituting relation R, and the same is true for partial control. However R is fleshed out in detail – for which there are several different theoretical options⁵ – it is obvious that R, and with it control, will not extend to all consequences of our choices, no matter how remote they are. It is even possible that control does not extend to results *at all*, and that R and control only extend to intentional bodily movements. Of course, it is also possible that there is *nothing* at all which stands in relation R to any agent, and, consequently, that nobody has control over anything.

The question of which view about the scope of control is correct depends on how the control-constituting relation R is to be understood. Even so, the purely structural points just made already show that premise (2) and the arguments given in its favour rely on a distorted view of control. Recall that Moore and Nagel base their defence of premise (2) on the assumption that if results are not within the scope of control, then nothing is. This conditional, however, has just shown to be false: even if results are not within the scope of control, it is nevertheless possible that things causally closer to the agent are (intentional bodily movements being a case in point).

7. In passing, it is worth noting that Moore's own way of setting up the matter also commits him to accepting the conceptual framework just introduced. In a passage from his 2009 book *Causation and Responsibility*, Moore takes what he refers to as a 'compatibilist' notion of control to be characterized by the thought that in order for there to be control over some result,

[w]e do *not* need to control *every* factor making [a] result possible [...]. (Moore 2009, 25, emphasis added)

⁴ This notion of partial control can be straightforwardly modified such that partial control comes in degrees: the more of those factors which are causally relevant for the occurrence of some X stand in relation R to some agent A, the higher the degree of partial control which A has over X.

⁵ Such as e.g. some version of agent-causality, or the notion of reasons-responsiveness as worked out in detail by Fischer/Ravizza (1998).

If what he calls an ‘incompatibilist’ notion of control is presupposed, by contrast, then

we control some results only when we can make causally efficacious choices about *every* factor that could cause or prevent this result. (Moore 2009, 24–25)

In a passage from his 1994 paper, he ascribes to incompatibilists essentially the same idea, using the notion of control instead of referring to the ability to make causally efficacious choices:

Incompatibilists have long sought to show that we must be in control over *all* factors that cause our choices in order for us to be responsible for both those choices and the wrongdoings such choices initiate. (Moore 1994, 257, emphasis added)

In order to avoid either conceptual circularity or a regress, the notion of control invoked in the *definiens* of these two definitions has to be different from the notions of ‘incompatibilist’ and ‘compatibilist’ control which they are, respectively, meant to define. As Moore sets up the matter, whether there is some sort of control – be it, in his terminology, a ‘compatibilist’ or an ‘incompatibilist’ one – over a result X, depends on whether we have control in some *other* sense over all or some of the factors leading to X. ‘Incompatibilists’ in Moore’s sense thus believe that in order to control anything at all, we need to be in control over all factors influencing it. ‘Compatibilists’ deny this.

‘Compatibilist’ control in Moore’s sense is partial control as introduced above. In order to define any of these two concepts, one needs to bring into play another, more basic notion of control: there is no partial control in the sense introduced above, and no ‘compatibilist’ control in the sense of Moore, without some more basic sort of control with regard to which partial control can meaningfully be defined. Apart from the fact that the above definitions require this to be so, the point can also informally be put as follows: when one starts thinking about results of actions as events or states of affairs, the occurrence of which is caused by a number of factors, then responsibility and control have not even begun to enter the picture. In order for control to be involved in some element in the causal chain of events, control has to enter the story at some earlier point.

Clearly, what Moore labels ‘incompatibilism’ is a highly implausible view, and there is much reason to reject it. A position which is ‘compatibilist’ in the nonstandard sense introduced by Moore, however, is open for both an incompatibilist and a compatibilist understanding of control, when these terms are used in the way which is standard in the debate about free will and responsibility. As already indicated above, the control-constituting relation R can be spelled out in

a number of different ways – and, in particular, such that relation R's holding is either compatible, or incompatible with determinism. In the former case, one's understanding of control (and, as a consequence, of partial control) is compatibilist. In the latter case, it is incompatibilist. It thus turns out that the contrast between compatibilism and incompatibilism is misconceived when it is seen as concerning the question of whether there is control over all, or only over some of the factors leading to some result.

8. The distinction between control and partial control which has been introduced in § 6 above has provided a basis for a criticism of premise (2) and, with it, of the thought that the problem of moral luck collapses into issues about free will. Given that control and partial control can (and, indeed, must) be distinguished conceptually, premise (2) is refuted, and so is the idea that, as far as control is concerned, there are only two options – either we control nothing at all, or all of those things which are potentially relevant for blameworthiness are within the scope of control.

If our understanding of control should, as has been argued above, be more nuanced, then the same is true of our understanding of the theoretical options with regard to the problem of moral luck. In order to assess whether certain elements of our moral practice are not in line with what is required by the control principle, relying exclusively on considerations about control does not suffice. What, if anything, is ruled out by the control principle depends both on what is within the scope of control *and* on what is a legitimate basis for moral assessment. Hence, in order to be able to assess how the distinction between control and partial control bears upon the appropriate understanding of the problem of moral luck, two sorts of considerations have to be brought together. For one thing, a more worked-out conception of control is required. In particular, the extent to which there is a tension between our practice and the control principle will depend on whether (and if so, why) control extends only to things causally close to the agent, such as intentional bodily movements, or whether (and if so, why) it also covers results. For another thing, at least the first steps towards a normative theory of blame need to be undertaken. In particular, it is of crucial importance whether blameworthiness depends only on things within the scope of control, or whether blameworthiness can also depend on what happens within the scope of partial control. One needs, in other words, to know whether control, or rather partial control, is the relevant notion figuring in the control principle.⁶

⁶ Note that this would be different if the scope of control is either void or else such that none of

Suppose that it turned out that control extends to results, or that blameworthiness attaches to partial control (or both). Then there would not even be the appearance of a fundamental tension between our moral practice and the control principle. Furthermore, *wide control* (as introduced above in § 5) would be the appropriate way of dealing with the problem of moral luck.

Correspondingly, there is only a tension between the control principle and our moral practice if the scope of control is restricted to intentional bodily movements (or something else causally close to the agent), and if, at the same time, blameworthiness is only dependent on factors within our control. Then certain elements of our moral practice would seem to violate what is implied by the control principle. To the extent to which this is so, both *reinterpretation* and *revision* are conceivable strategies. Note that also a combination of these two options offers an attractive perspective: while certain reactions might indeed be out of place and should be abandoned, others might be appropriate but misinterpreted if understood as being based on the idea that blameworthiness depends on results.

9. The task of developing reasonably worked-out conceptions of both control and of (the normative grounds of) blameworthiness cannot be accomplished in this paper. Even so, here shall be told at least the beginning of a plausible theoretical story supporting the view that the scope of control and the scope of what blameworthiness depends on coincide.

As a number of authors have observed, it is a worthwhile project to spell out the notion of control by fleshing out the thought that those sorts of things which are within the scope of what agents can control are exactly those things with re-

the relevant items are beyond what agents can control (and that, in particular, results are within the scope of control). Then the control principle either implies that none of our moral assessments are acceptable, or it does not constrain them at all, and other sorts of considerations can be brought into play. This is indeed why Moore believes that he can do without both a worked-out conception of control and a normative theory of blame in dealing with the problem of moral luck. The view that the blameworthiness of an agent is increased, *ceteris paribus*, when the agent's action has caused some harm could then be backed up with reference to the way we react in situations of the kind in question. This is precisely what Moore does using what he calls an "experiential argument" (Moore 2009, 29), which allegedly allows to infer that

we *are* more blameworthy when we cause some evil, than if we merely try to cause it, or unreasonably risk it (Moore 2009, 30)

from the fact that

we experience greater guilt when we have caused some harm that we either tried to cause, or unreasonably risked, than we experience when we have been equally culpable but we have not caused such a harm (Moore 2009, 30).

spect to which agents can be said to be responsive to reasons. The *prima facie* attractiveness of this idea is supported by the following pair of considerations: for one, being an agent does seem to be closely related to being able to recognize and to being guided by reasons for action. For another, the scope of control is the sphere of agency.

Given that we consider only agents to be potential candidates for appropriate blame, there is reason to believe that blame is, in a sense to be worked out, essentially tied to agency. This observation at least suggests that blameworthiness will have something to do with what is, or could have been, the case with regard to things within the scope of control. This idea gains additional support by the observation that retrospective blame can be understood as something like a temporary shifted correlate of demands – i.e. a correlate of speech acts which are meant to influence agents by giving them, or pointing out to them, certain reasons.⁷

There is, thus, reason to believe that the scope of control and the scope of what our blameworthiness depends on coincide. Moreover, it is not implausible to assume that actions, and not results, are the sorts of things with regard to which agents are responsive to reasons and which, consequently, are relevant for our (degree of) blameworthiness.

This is, to be sure, nothing more than a very brief hint at an argumentative strategy supporting the view that blameworthiness and control – as opposed to partial control – are intimately connected, and that the scope of both includes actions, and not results. If this strategy can be successfully developed, then the tension at the heart of the problem of moral luck is that our moral practice

⁷ It is instructive to consider the case of criminal law and criminal responsibility. With regard to criminal law, Stephen Morse points out that

a crucial aim of the criminal law is to guide and thus to prevent certain *actions*, because only actions can culpably produce harms. (Morse 2004, 367)

Given that, according to Morse,

legal rules [...] guide actions primarily because they provide an agent with good moral or prudential reasons for forbearance and action (Morse 2004, 368)

and also given that

intentional action or forbearance is the only aspect of the human condition that is fully ‘up to us’, that is fully within our control, and that can be fully guided by and produced by our reason (Morse 2004, 368)

(the degree of) criminal responsibility should depend on actions, and not on their results, since results cannot be controlled by reason(s) in the same way as actions. These ideas cannot, however, be straightforwardly transferred to the case of moral blameworthiness, since a purely instrumentalist interpretation of moral blame is, at least initially, not as plausible as it is in the case of criminal law.

sometimes seems to rest on a view which cannot be maintained. Attempts to re-interpret our reactions in accordance with the idea that upholding our practice does not commit us to the view that blameworthiness depends on results, or attempts to revise our practice to the extent to which it violates the control principle, or a combination of these two strategies, are the only theoretically feasible options.⁸

References

- Enoch, David & Marmor, Andrei (2007). The Case Against Moral Luck. *Law and Philosophy* 26, 405–436.
- Fischer, John M. & Ravizza, Marc (1998). *Responsibility and Control. A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Moore, Michael S. (1994). The Independent Moral Significance of Wrongdoing. *Journal of Contemporary Legal Issues* 5, 237–281.
- Moore, Michael S. (2009). *Causation and Responsibility. An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press.
- Morse, Stephen J. (2004). Reason, Results, and Criminal Responsibility. *University of Illinois Law Review*, 363–444.
- Nagel, Thomas (1976). Moral Luck. Quoted from the revised version reprinted in Thomas Nagel, *Mortal Questions*. Cambridge: Cambridge University Press 1979, 24–38.
- Nelkin, Dana (2008). Moral Luck. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/moral-luck/>.
- Strawson, Peter F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48, 187–211.
- Williams, Bernard (1976). Moral Luck. Quoted from the revised version reprinted in Bernard Williams, *Moral Luck*, Cambridge: Cambridge University Press 1981, 20–39.

⁸ I have presented versions of this paper at workshops and colloquia in Aachen, Berlin, and Potsdam, and I have profited a great deal from the discussions which took place on these occasions. I also received very helpful comments on a draft by Anne Burkard, Simon Gaus, Jan Gertken, Benjamin Kiesewetter, Leo Menges, Christian Neuhäuser and Eduardo Rivera Lopez.