Agency and Mistakes

Santiago Amaya
Universidad de los Andes

Mistakes are part of our life as agents. Although we tend to be shy about them, preferring sometimes not to talk about them, everyone recognizes having made a mistake at some point or another. Philosophers are no different in this regard. In theorizing about agency, we tend to emphasize the successful cases. Even when we recognize our imperfections, we mostly focus on those that only derivatively show up in what we do.

Mistakes, however, are more than a pervasive feature of our lives. Even if one worries about making them, there is still something positive about them. Not only, as the saying goes, can we learn a lot from our mistakes; we can also learn a lot about what kind of agents we are by thinking about the mistakes we make. There is, in fact, as we shall see, reason to think that mistakes are built into the capacities that make us agents.

In this chapter, I discuss the relationship between agency and mistakes. I begin with some clarifications that show why this relationship matters (sections 1-2). Then, I offer a quick survey of the mistakes often discussed by philosophers of action, noting the tendency to focus on those explained by errors in some of our attitudes (section 3). With this, I discuss the category of performance mistakes and some strategies that have been deployed to downplay their importance (sections 4-5). I conclude with some brief remarks about the sources of our fallibility as agents (section 6).

1. Standards of agency

Human creatures have a distinct set of abilities in virtue of which we qualify as agents. We can, for instance, form desires and intentions. We can normally figure out which courses of action are better to satisfy those desires and are consistent with our intentions. And we can settle on those courses of action or renounce from embarking on others, as the case might be. Philosophical accounts of agency emphasize different abilities in identifying what is constitutive of human agency.

Humans, of course, are not the only agents out there. The animal kingdom, as some have noted (see, for instance, Andrews, 2013; Burge, 2009), abounds with instances of non-human agency: creatures coordinating their behaviors to achieve goals in ways that are attributable to them. For present purposes, however, we shall restrict ourselves to what might be described as *distinctively human* agency. That is, we shall focus on instances of agency that require the exercise of abilities that seem exclusively within the purview of humans.

Even then, there is room for disagreement. For many theorists, interest in human agency reflects foremost an interest on *intentional actions*: roughly, the things that you do in the execution of your intentions (Brand, 1984; Mele, 1992). Others have thought that our capacity to form intentions and plan ahead of time has a certain priority here, so their focus

has mostly been on our capacities for *planning agency* (Bratman, 1987). Finally, some take their interest in agency as derivative on questions related to autonomy and freedom. Hence, they have argued that in theorizing about human agency the spot should not shine equally on all intentional actions but, instead, on *full-blooded instances* of them (Velleman, 2000).

Being an agent, however, requires more than the possession of these abilities, regardless of which of them one's preferred theory identifies as central. It is a matter as well of shaping what we do by virtue of their exercise. This last bit is important. Desiring good things or having true beliefs about how to achieve them, forming coherent plans, and making decisions that accord to them, are certainly remarkable things. But, unless you can translate those attitudes into behavior and structure what you do accordingly, you won't be an agent. You might, at best, be a good thinker.

Thus, a widely agreed feature of actions is that, by virtue of being brought about by agents, they are open to certain forms of evaluation. This is true whether we speak about minimal or distinctive human agency, of intentional, planning, or full-blooded agency (for interesting discussion, see Shepherd, 2021, ch. 6). In brief, being capable of behaving in certain ways creates *reasonable expectations* that one would behave in those ways, at least in some circumstances. In doing so, it gives rise to standards by which one's doings can be evaluated.

Naturally, different standards matter for different kinds of agents. In the case of minimal forms of agency, for instance, actions are typically evaluated in terms of survival, reproductive success, and the like. With respect to our actions, standards of evaluation vary widely with circumstances and interests. Although some are specific to a domain, a few seem to be domain general. In principle, any of one's actions can be judged in terms of how prudent, moral, or rational it is.

2. Which standards?

In speaking of standards of evaluation, we should be careful to distinguish two different notions. By not doing it we might end up with a picture that has a hard time accommodating some of our mistakes (for one crisp way of bringing up these difficulties, see Lavin, 2004, sect 8) There are, first, what have been called *constitutive standards*. These are standards that any behavior would have to meet to be considered an exercise of our agency. In addition, there are what we might call *standards for success*: those that exercises of agency need to meet to be considered successful, that is, to be successful *qua* actions.

Most theorists agree that the constitutive standards of human agency are tied to certain expectations of rationality, which is why human agency is often glossed as *rational* agency. In short, any behavior that qualifies as a human action ought to have some sort of rational structure. For different kind of theorists, constitution will mean different things: being caused in certain ways, being responsive to certain norms, having some distinctive kind of explanation, etc. And rationality will be tied to different set of norms. But, ultimately, the underlying thought is that human actions are somewhat rational responses (not necessarily the most rational responses) to the circumstances we occupy.

Undoubtedly, more variation can be found with respect to the expectations by which actions get evaluated for success. Among other things, what counts as successful varies from one domain to another. But, arguably, a minimum success standard for intentional action, and for more sophisticated forms of agency that depend on it, is that it satisfies the intentions that give rise to it (see Amaya, 2017; Radoilska, 2013). A successful action, on this view, is one whether the agent achieves what she originally intended with it, given the attitudes that the agent had at the time.

Now, constitutive and success standards, even those that are domain general, can evidently come apart. Cases of *causal deviance*, such as those illustrated by Donald Davidson's (1978) famous climber, are a good example of how success standards could be met, except that the conduct in question fails to meet constitutive standards for agency. In such cases the behavior matches the intentions of the agent, so in that respect what the agent does would seem a success. But to the extent that it is not the product of the rationality of the agent, it falls short of exhibiting the structure that, according to many theorists, is required of genuine action.

Mistakes show how these standards can come apart, but here the dissociation goes in the opposite direction. At least, this is true of the kind of mistakes that matter in thinking about our agency. When we make a mistake, what we do can be described as meeting the constitutive standards of agency, say, by exhibiting the rationality characteristic of actions But it fails to meet some reasonable standard of success. Thus, although at the time we behave in ways that are somewhat reasonable, at least reasonable enough to be considered actions, what we do falls below some reasonable standard. At the very minimum, we fail to act as intended.

Consider, in this regard, two common ways of rejecting imputations of a mistake. First, one can disavow the imputation by questioning whether one's actions actually violated a reasonable standard of success. Expressions such as "I didn't know," "It was too hard for me to do it," or even "I did mean to do that," often mark this sort of response. Alternatively, instead of questioning whether a mistake was made, one can reject the imputation by questioning the status of what one did as an exercise of one's agency. "It was an accident," "I could not help it," "I was just bad luck," are expressions typically used to indicate this second way of distancing.[1]

We can now say more precisely why our mistakes are a window into the kind of agents that we are. By dissociating constitutive and success standards, reflection on our mistakes

[1] The distinction is similar, but not the same, as the one that Peter Strawson (1962) draws between excuses and exemptions. Exemptions (e.g., appealing to insanity) call into question the status of the person as a morally responsible agent altogether (Watson 1987, Wallace 1994). In the distinction drawn above, instead, the exclusion is supposed to be piecemeal: the person might generally be considered an agent but not with the respect to the behavior in question. Excuses, on the other hand, need not be full exculpations. "I did not mean to do that" might show that I didn't act of ill will, but bad will is not necessary for blame attributions. For discussion, see Amaya and Doris (2014)

(which are they and why we make them) require us to think what are the conditions that exercises of our agency need to meet to be such, as opposed to those that we simply find desirable that our actions meet. Doing so, requires not just that we specify clearly what those standards are, but also that we ask ourselves how much imperfection our conception of agency can tolerate. Ultimately, the size of the gap between constitutive and success standards, especially those that are domain general, will reveal how accepting we truly are of our mistakes.

3. Which mistakes?

Some philosophers might not worry whether their accounts of agency are actually instantiated by imperfect human beings (see, for instance, Velleman, 2004 and Ford 2011). *Idealization* is somewhat widespread in this corner of philosophy. But many of them aim at providing theories that seek to recognize our human fallibility. So, even if explicit theorization of mistakes is rare, mention of them tends to show up in discussions about human agency.

Yet, it is important to observe which kinds of mistakes normally get recognized and what role they are typically called to play in the construction of these accounts. Often the mistakes discussed can be traced to problematic features of the attitudes behind our actions, the goals we desire, how we think they can be achieved, etc., as opposed to being non-derivative mistakes of our agency. As a consequence, most theorizing in this area tends to overlook the extent of our imperfections as agents.[2]

It is easy to recognize this trend once one's attention is drawn to it. In his seminal essay on *Agency*, for example, Davidson (1971) sought to identify what sets our mistakes apart from unintended things that merely happen to us. Among the examples discussed, he mentions the man who unintentionally spills his coffee, the agent who boards the wrong plane, the prince who stabs an innocent man, and similar cases. In all of them, the source of the mistake is the same: the person thought the cup had tea (not coffee), he believed the plane was heading to London, England (not London, Ontario), and Hamlet didn't know Polonius was behind the tapestry.

As Davidson knew well, mistakes originating in our conative, as opposed to our doxastic, attitudes are also possible and part of our life. These have also received a fair share of philosophical attention. In their groundbreaking discussions of autonomous agency, for instance, Harry Frankfurt (1971) and Gary Watson (1975) brought to light cases of agents alienated from their desires, asking to what extent their actions could be considered free and attributable to them. Less dramatic, but equally problematic perhaps, are David Gauthier's (1986) examples of agents that form preferences lacking any real information about the options presented to them.

---

[2] Exceptions to this generalized trend can be found in Mele (2006), Peabody (2005), and Frost (2018), but J.L. Austin's (1957) "Plea for Excuses" is a predecessor worth mentioning too.

In addition, philosophers have also discussed cases of attitudes that do not fit well with each other. For this lack of alignment might also be conducive to making mistakes. Discussion about weakness of will is good example here. As Amelie Rorty (1980) noticed, episodes of weakness of will can involve a multiplicity of different conflicts. They could happen, as Alfred Mele (1988) has argued, when we form intentions that run contrary to the agent's considered judgements. Or, as Richard Holton (1999) described them, the conflicts can be among our current intentions and resolutions to overcome temptation that we made in the past.[3]

It is worth noting the pattern behind these well-known discussions. Although the examples brought up there involve recognizable mistakes, they are ones that only derivatively we make as agents. In them, agents fail to succeed in some significant respects, say, failing to obtain what they "really" want or what they would have intended had they controlled themselves. But they fail at it because they antecedently have problematic beliefs, they desire things that they do not really want or they think they shouldn't want, or because the relevant attitudes behind their actions are not properly aligned with each other. The standards by which these actions are mistakes are the standards of correctness for those attitudes.

Unfortunately, this leaves too many of our imperfections out. As we mentioned at the outset, human agency is not exercised simply by having attitudes of a certain sort (say, those that track true and good things) or that bear certain kind of relations (of consistency, for instance) with each other. In order to be considered agents, we also need to effectively translate those attitudes into action. And here, in the process of making these transitions, the possibility of making a variety of mistakes of a different sort opens up. Despite having attitudes that are not, in principle, questionable, we can end up doing things that fall short of some reasonable expectations of success.

Let me make the overall pattern more perspicuous by distinguish two kinds of mistakes. There are, first, mistakes that can be traced to the attitudes that lie behind the actions that we undertake. In these cases, the attitudes themselves fall below certain standards applicable to them or they instantiate patterns that otherwise violate some standard of correctness. These are the mistakes that mostly figure in the philosophical literature. But there are, in addition to them, *performance mistakes*. These are ones that we make, not as believers, desirers, etc., but as creatures who can translate their beliefs, desires, and so forth into actions. In so far as these do not necessarily impugn our attitudes, they are mistakes as judged by standards of success independent of whichever measure of correctness we use to evaluate our attitudes. Because of this, they illustrate how our fallibility extends beyond our problematic attitudes.

4. Performance mistakes

[3] Not all treatments of weakness of will follow this general pattern. There are ways of thinking about akratic breaks, where these involve a person acting against her considered judgment or prior resolution but where her action is the product of a habit, not so much of a conflict among her preferences (see Silver 2019 and Mylopolous & Pacherie 2020).

To get a sense of how this happens, consider the following quotation from Michael Bratman:

> Suppose that this morning I formed the intention to have a milkshake at lunch, lunchtime arrives, my intention remains, and nothing unexpected happens. In such a case, I do not normally need to tote up the pros and cons concerning milkshake drinking. Rather, in the normal course of events I will simply proceed to execute (or, anyway, try to execute) my intention and order a milkshake. My intention will not merely influence my conduct, it will control it (1987, p.16)

The passage is meant to introduce Bratman's influential idea that intentions involve a characteristic commitment, something that sets them apart from other attitudes. But what he says in it is *literally* false. Intentions, we can agree, control conduct in ways in which, say, desires don't. They involve a commitment to act when time comes. Yet, living up this commitment requires more than forming an intention and waiting for that time to arrive.

Imagine some mundane ways in which Bratman's story might unfold instead. You decide to have a milkshake, the intention remains, and so forth. But, as soon as you get to lunch, you see in the menu a combo that includes a soda and that looks like a good deal. So, you go ahead and order it. It is not as though you changed your mind and desist from having the milkshake. Simply, you ordered something else without even *remembering* your prior intention. Alternatively, at noon you leave the office to get lunch, while still thinking about the paper you were writing. But *out of habit* you end up walking to the place with the salad bar where you normally go, instead of walking to burger place where you had earlier planned to go. Only after you finish your salad and the mineral water you order with it, you realized you never had the milkshake.

These stories should sound familiar, as they illustrate a common kind of performance mistake: action slips. Each, however, instantiates a slightly different pattern (Norman, 1981; Reason, 1984). Sometimes, as in the first story, you form an intention to do something later and plan accordingly, but the thought of your early plans simply does not come to mind when the time is right. So, you end up doing something that would otherwise makes sense to do then (Amaya, 2013). Or, as in the second story, being engrossed in some mental task, a habitual routine kicks in and you end up doing something that predictably would not allow you to do what you intended first (Amaya, 2020).

Slips, to be sure, tend to result in minor inconveniences (although some slips are genuinely catastrophic, see Amaya & Doris 2014). And normally they do not occur because of some problematic attitude that you need to change. Certainly, it was not wrong for you to find the lunch combo appealing; the habit of going to the salad place is definitely a good one to have. In any other occasion acting on that desire or following that habit would have been an acceptable thing to do. However, on this occasion, by acting on them, you ended up acting contrary to the intentions and plans that you had. By doing so, you ended up making a mistake.

The point can be put in more general terms. Being an agent, let's agree, requires the ability to form intentions and plans for executing them. But having that ability, even exercising it,

only makes you a good planner. This is what Bratman fails to note in the passage above. To be an agent, you also have to exercise the power of bringing to mind those intentions and plans in the midst of doing other things and with potential distractors in sight. And here, in exercising these powers, mistakes can occur. If you fail to keep track of where you are with respect to your overall plans, whatever you intended to do might never materialize.

Obviously, slips are only one among many possible illustrations here. Life is full of episodes of forgetfulness, distractions, and the like, to be all fitted in one and the same category. So, in addition to slipping, performance mistakes can happen in all sorts of ways: in the ways in which our desires and judgments connect with our decision-making processes, in the way in which what we know informs how we deliberate, etc. When these disconnects happen our actions will fail to reflect the relevant attitudes. So, even if they have all the right desires, beliefs, etc., and even if these might be rightly aligned with each other, they might fail to make a difference in what we do.

5. Attitude-based explanations

The term "performance mistake" probably resonates with G.E.M. Anscombe's followers. In a well-known passage of *Intention*, she marks the distinction drawn above (at least, a version of it) by contrasting mistakes in judgments and those that occur in performance. Alas, Anscombe did not spent much time explaining the contrast, nor did she offer much by way of discussion of the latter sort of mistakes. Her attitude, instead, was to casually dismissed them the "rare exception." (Anscombe, 1963, par. 32).

Other philosophers may have a similar attitude: they recognize that performance mistakes can happen, but they treat them as oddities in the life of agents. This actually might explain Bratman's oversight above or, at least, why he hedged at various points of the passage using the word "normally." The truth, however, is that these mistakes are more common than one might initially think, definitely at least as common, say, as the episodes of alienation or weakness of will over which so much ink has been spilled. If anything, judging by the recurrence of first-person reports, episodes of misremembering, getting distracted, and they like, should not count as unexpected (Jónsdóttir et al 2007; Reason & Mycielska, 1982).

Interestingly, even some who have recognized the pervasiveness of performance mistakes have fallen short of pressing the idea further. Their reaction, instead, has been mostly to try to explain them away, either by reducing our performance mistakes to prior errors in our attitudes, or placing them altogether outside the domain of our agency. If idealization is sometimes widespread among philosophers of action, perfectionism is rampant. The prototypical actions that our models of agency are meant to explain, almost as a matter of rule, are successful actions.

To see how these attitude-based explanations work, let us return to slips. Without a doubt, the grandfather here is Sigmund Freud, who was perhaps the first to recognize the pervasiveness and psychological significance of them (1901, 1915). In his early writing, he complained about what he referred to as "psycho-physiological" explanations of the slip, in brief, explanations of them as glitches in the cerebral machinery behind our conduct.

Instead, for him, slips were a window into the deeper mechanisms of the human mind. In particular, they were mistakes that resulted from the fact that people regularly held in the minds emotionally threatening motives.

Nowadays, the influence of Freud and his psychodynamic approach is less gripping than it used to be (although some of it survives in the writing of some philosophers; see, for instance, Arpaly, 2003 pp 159-162, and Velleman, 2000, intro). But this has often meant, not a retreat from attitude-based explanations, but a refashioning of them in the guise of new psychological models of the human mind.[4] Just to take an example, Roy Sorensen (2011) argues that popular dual-processing models in cognitive psychology underwrite a distinction between deep and shallow attitudes. According to him, slips are actions caused and rationalized by shallow beliefs and desires that happen to be out of line with our more reflective attitudes.

We cannot go here into a detailed criticism of these approaches. But it is worth noting how these explanations often look very much like *ad hoc* reconstructions. Although the attitudes invoked to explain the slips might make sense of the behavior of the person at the time, they often fail to make sense of the overall behavior of the person. Normally, there are simpler and more comprehensive available explanations around the corner, explanations that are more transparent of the psychology behind the mistake.

To illustrate, return one last time to the alternative to Bratman's story. Let's us accept that it is possible that you developed a sudden desire for salad. And, if that were the case, that would explain why you ended up going to the salad bar instead of the burger place. But it's more plausible to say that what got you there was the habit of going for lunch to that place almost every day. That was, in fact, a habit that you had—you often went there for lunch there. It accords with your state of mind at the time—habits tend to capture behavior when one's mind is somewhere else. And it explains why what you end up doing that day, despite being a mistake, was not something out of the ordinary.

Unsurprisingly, this kind of criticism is one that detractors of Freud have repeatedly pointed out (see especially Grünbaum, 1984). The evidential basis for attributing agents the specific repressed motives that psychodynamic explanations involve is often thin. Except for the isolated episode of the incident they do not show up anywhere else in the life of the person. If anything, what seems to drive Freud's explanations is almost entirely his prior commitment to the theory.

---

[4] By contrast, the psycho-physiological approach to slips that Freud denounced mostly faded away. Still, off-hand remarks echoing these approaches can be sporadically found in the literature. Back in the day, for instance, Hilary Putnam (1975, p. 372) compared our thinking processes to the operations of Turing machine; he then went on to compare verbal slips with accidents in the printing device of those machines. Likewise, when it was objected that slips were not explainable from the intentional stance, Daniel Dennett dispatched the objection declaring slips "either malfunctions or outcomes of misdesign" (1982, p. 63).

In the end, it is possible that something similar happens here: what drives attitude-based approaches to our performance mistakes are other prior philosophical commitments, plausibly, a commitment that all explanations of action need to have a common form (for two very different expression of this commitment, see Smith, 1998; Thompson, 2008). Certainly, explaining how performance mistakes happen and why they still count as exercises of our agency requires accepting a more pluralist approach to the standard of rationality that are constitutive of our actions (for a defense of this pluralism, see Amaya, 2020). It requires, for instance, understand that habitual actions might meet the constitutive requirements of rational actions in ways that differ from actions that rationalized by beliefs and desires.

6. Facing imperfection

We began with the idea that, by dissociating constitutive and success standards, reflection on our mistakes is a window into the nature of human agency. As we have seen, most philosophical reflection has been concerned with mistakes that derive from errors in the formation of our attitudes or on how these attitudes fit together. So, in the end, most of it has focused only on conducts that qualify as mistakes in the sense that the attitudes behind them fall short of correctness standards for them.

It is, no doubt, important to have an appreciation of this kind of mistakes. But concentrating on them runs the risk of painting an inaccurate picture of ourselves, one that locates that source of our imperfections as agents in the wrong place. Focusing on performance mistakes, by contrast, paints a different picture. To the extent we not only intend, plan, and decide what to do, but that it is also upon us to translate our intentions, plans, and decisions into conduct, we do not make all of our mistakes as believers, deciders, etc. As revealed by our performance mistakes, we also have limitations having to do directly, not derivatively, with our agency.

If true, these considerations recommend that we endorse a *kind of fallibilism* with respect to our agency. According to it, the possibility of making mistakes is built into whichever powers make us agents. Ultimately the thought here is that the gap between the constitutive standards of our agency and our standards of success is not entirely accounted for by the limitations we have in other domains of our rational lives.

References

Amaya, S. (2013). Slips. *Noûs*, *47*(3), 559–576. https://doi.org/10.1142/9789814569057_0006

Amaya, S. (2017). Two kinds of intentions: a new defense of the Simple View. *Philosophical Studies*, *175*(7), 1767–1786. https://doi.org/10.1007/s11098-017-0934-1

Amaya, S. (2020). Out of habit. *Synthese*. https://doi.org/10.1007/s11229-020-02780-3

Amaya, S., & Doris, J. M. (2014). No Excuses: Performance Mistakes in Morality. In J. Clausen & N. Levy (Eds.), *Springer Handbook of Neuroethics* (pp. 253–272). Netherlands: Dordrecht.

Andrews, K. (2013). Ape autonomy? Social norms and moral agency in other species. In K.

Petrus & M. Wild (Eds.), *Animal Minds and Animal Ethics*. Bielefeld: Transcript.
Anscombe, G. E. M. (1963). *Intention*. Harvard University Press.
Arpaly, N. (2003). *Unprincipled Virtue*. New York: Oxford University Press.
Austin, M. A. (1957). A Plea for Excuses. *Proceedings of the Aristotelian Society*, *57*, 1–30.
Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. MIT Press.
Bratman, M. E. (1987). *Intentions, plans and practical reasoning.* Cambridge: Harvard University Press.
Burge, T. (2009). Primitive agency and natural norms. *Philosophy and Phenomenological Research*, *70*(2), 251–278.
Davidson, D. (1971). Agency. In R. W. Binkley, R. N. Bronaugh, & A. Marras (Eds.), *Agent, Action, and Reason* (University, pp. 1–37).
Davidson, D. (1978). Intending. In *Philosophy of history and action* (pp. 41–60). Springer, Dordrecht.
Dennett, D. C. (1982). Making sense of ourselves. In J. I. Biro & R. W. Shahan (Eds.), *Mind, Brain, and Function* (pp. 63–81). University of Oklahoma Press: Norman.
Ford, A. (2011). Action and Generality. In J. Hornsby, F. Stoutland, & A. Ford (Eds.), *Essays on Anscombe's Intention* (pp. 76–104). Harvard University Press.
Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5–20.
Freud, S. (1901). *The Psychopathology of Everyday Life*. London: Hogarth Press.
Freud, S. (1915). *Introductory Lectures on Psychoanalysis*. London: Hogarth Press.
Frost, K. (2018). Basic mistakes in performance. *Proceedings of the XXIII World Congress of Philosophy 44*, 17–21.
Gauthier, D. (1986). *Morals by agreement*. Oxford University Press on Demand.
Grünbaum, A. (1984). *The foundations of psychoanalysis: A philosophical critique*. University of California Press.
Holton, R. (1999). Intention and weakness of will. *The Journal of Philosophy*, *96*(5), 241–262.
Jónsdóttir, M., Adólfsdóttir, S., Cortez, R. D., Gunnarsdóttir, M., & Gústafsdóttir, Á. H. (2007). A diary study of action slips in healthy individuals. *The Clinical Neuropsychologist*, *21*, 875–883.
Lavin, D. (2004). Practical reason and the possibility or error. *Ethics*, *114*(3), 424–457. https://doi.org/10.1111/1467-9213.00296
Mele, A. R. (1988). Irrationality: a précis. *Philosophical Psychology*, *1*(2), 171–177.
Mele, A. R. (1992). *Springs of action: Understanding intentional behaviour*. Oxford University Press.
Mele, A. R. (2006). Practical Mistakes and Intentional Actions. *American Philosophical Quarterly*, *43*(3), 249–260.
Mylopoulos, M., & Pacherie, E. (2020). Self-control as hybrid skill. In A. R. Mele (Ed.), *Surrounding Self-Control* (pp. 81–100). Oxford University Press.
Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*(1), 1–15.
Peabody, K. (2005). Trying slips: Can Davidson and Hornsby account for mistakes and slips? *Philosophia*, *33*, 173–216.
Putnam, H. (1975). Minds and machines. In *Mind, language and reality.Philosophical papers* (pp. 362–385). Cambridge, MA: Cambridge University Press.
Radoilska, L. (2013). Addiction and weakness of will: An integrated account. In *Addiction*

and Weakness of Will (pp. 119–139). Oxford University Press.
Reason, J. (1984). Lapses of attention in everyday life. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 515–549). New York: Academic Press.
Reason, J., & Mycielska, K. (1982). *Absent-minded? The psychology of mental lapses and everyday errors*. Englewood Cliffs, N.J: Prentice-Hall.
Rorty, A. O. (1980). Where does the akratic break take place? *Australasian Journal of Philosophy*, *58*(4), 333–346. https://doi.org/10.1080/00048408012341341
Shepherd, J. (2021). The Shape of Agency. In *The Shape of Agency: Control, Action, Knowledge* (pp. 57–91). Oxford University Press.
Silver, K. (2019). Habitual weakness. *Thought: A Journal of Philosophy*, *8*(4), 270–277.

Smith, M. (1998). The possibility of philosophy of action. In *Human action, deliberation and causation* (pp. 17–41). Springer, Dordrecht.
Sorensen, R. (2011). What lies behind misspeaking? *American Philosophical Quarterly*, *48*(4), 399–409.
Strawson, P. (1962) Freedom and Resentment. *Proceedings of the British Academy* 48:187–211.
Thompson, M. (2008). *Life and action*. Harvard University Press.
Velleman, D. (2000). Introduction. In *The Possibility of Practical Reason* (pp. 1–31). Oxford: Oxford University Press.
Velleman, D. (2004). Précis of The Possibility of Practical Reason. *Philosophical Studies*, *121*, 225–238.
Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge Mass.: Harvard University Press.
Watson, G. (1975). Free agency. *The Journal of Philosophy*, *72*(8), 205–220.
Watson, G. 1987. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Schoeman, 256–86. Cambridge: Cambridge University Press.