

# Out of Habit

Santiago Amaya  
Universidad de los Andes

Accounts of intentional agency typically assign beliefs a central role in the genesis of action. According to them, if someone acts with an intention, the person's action is guided by her beliefs. The beliefs guide the action by directing the agent towards ways of achieving what she intends. In this respect, they help rationalize it. They explain the action by revealing the connection the person saw between what she did and what she intended with it.<sup>1</sup>

Slips are a challenge to this view. When you slip, you act with an intention, but you inadvertently end up doing something else. For example, in early January you write the wrong year on a check or, right after moving to a new place, you find yourself driving to the old one. Surely, you are aware that the year ended—New Year's celebrations took place a few weeks ago—and you know that you just moved—you were actually coming back from returning the rental truck, but nevertheless, by acting *out of habit* you behave contrary to what you believe then.

It might be tempting to resist the challenge in two different ways. One is to claim that slips are not really instances of acting with an intention. Another is to try to reduce the slip to a doxastic mistake. Here, I argue both ways of fending off the challenge fail, because they wind up mischaracterizing the mistake. In the slip the agent acts with an intention. But she does not act guided by her beliefs about how to execute it.

---

<sup>1</sup> Versions of this idea can be found in Davidson (1980: 3-5, 85-86; 231-232; 266-267) and other causal approaches to agency (Audi 1994; Smith 2010; and Arpaly and Schroeder 2014). Even detractors of the causal theory have seen doxastic guidance as a principle constitutive of it (See Ruben 2003). The idea, however, goes beyond this type of approach. There are Kantian (Korsgaard 1997) and volitionist (Ginet 1990; Wallace 2001) versions of it too. Here, I discuss the challenge in the context of a causal theory but the challenge is meant to be broader than this.

Alternatively, one might try to discount slips on the grounds that the principle was never meant to apply to them. Yet, as we shall see, this response fails too. Slips have a *rational structure* of the sort that accounts of intentional agency are meant to capture. They exhibit robust and regular patterns, which speaks against them being fortuitous *glitches*. Thus, by helping dissociate habitual from other ways of acting intentionally, slips help understand how habits can rationalize what we do.

This last point is worth emphasizing. Philosophical accounts of habitual action tend to be psychologically shallow; psychologists typically contrast the habitual and the intentional with each other.<sup>2</sup> Thus, at present we have no theoretical account of how a psychology of intentions can support everyday habitual action. The discussion of slips here obviously does not amount to a full theory but it helps sketch how such an account can go.

The paper begins with a statement of the challenge, introduced with a real-life case that gets discussed throughout. Then, I proceed in two stages, discussing the ways of resisting the challenge mentioned above and using the conclusions reached at each stage to elaborate a view of the psychology behind the slip. Only in the end, I say how the results obtained can be generalized into a fuller theory of everyday habitual action. Developing that theory, as the case discussed here illustrates, requires that we think harder about the role that our history and our institutions play in shaping what actions become habitual for us.

## Oakland

---

<sup>2</sup> Although philosophers have said little about it, when habitual action has been discussed, it has often been presented as a counter-example to psychological conceptions of intentional agency (Pollard 2006; Di Nucci 2011. But see Douskos 2019 for a nice exception).

Psychologists interested in automatic goal-pursuit have theorized to a slightly larger degree about habits. Aarts & Dijksterhuis (2000), for example, define habits as goal-response associations that become active without *intent*. Wood & Neal (2007) conceptualize them as stimulus-response associations that structure behavior without mediation of *actual* goals.

While useful for capturing some forms of automaticity, these ways of thinking about habits are removed from interest in intentional agency.

On January 1st, 2009 police officer Johannes Mehserle shot Oscar Grant dead on a BART platform in Oakland, California. Mehserle didn't intend to shoot Grant. Yet, after warning him loud and clear that he would be tased if he continued resisting arrest, the officer pinned him down, drew the weapon, and pulled the trigger once. Tragically, although Mehserle meant to reach for the taser, he instead reached for his gun.<sup>3</sup>

From the perspective of the prosecution, the killing was intentional. They argue that it was unreasonable to suppose, as Mehserle claimed, that he had confused his gun with his taser. Although both weapons look alike, they feel very differently when held in your hands. Further, given that he was carrying them in opposite sides of his body, deploying each weapon required making very different physical movements. As a right-handed officer, he was carrying the gun in the right holster for a strong-side draw. He was carrying the taser in the left with the butt pointing forward in a cross-draw configuration.

The defense, on the other hand, claimed that the killing was accidental. Mehserle had neither intentionally used a deadly weapon, nor evinced disregard for the outcomes of his actions. The killing was just the consequence of the training he had received. After many hours of rehearsal with a gun and no more than a few hours of taser training, he had internalized the routine of deploying the weapon in his right holster when subduing a suspect. In fact, for them, the confusion was the only explanation consistent with his behavior before and after the shooting. Immediately after it happened, Mehserle brought his hands to his head and exclaimed with disbelief, "Oh shit, oh shit, I shot him."

The case was, predictably, plagued with controversy. The prosecution, for example, argued that Grant was actually trying to comply with the officers' orders when he was shot. Lying face down, his hands were just stuck under the weight of his body. Further, as the

---

<sup>3</sup> For discussion of the case, see *People vs. Mehserle*, #A130654, 206 Cal. App. 4<sup>th</sup> 1125, 142, Cal. Reprt. 3d 423 (2012). In what follows, I use the evidence and discussion documented in the court case as a guide to describe what happened that night in Oakland. Obviously, as with any other real-life case, the truth about Mehserle's mental states might turn out to be somewhat different. Also, in following the court case, I do not mean to deny that police shootings (in the US and elsewhere) sometimes get falsely presented as non-intentional. But exploring these larger issues concerning police brutality and their accountability for it is something beyond this essay.

District Attorney pointed out in his closing argument, the six cases of gun/taser confusions brought up by the defense involved a police officer carrying the weapons next to each other, not on opposite sides of his body.

This was probably not the best choice for the District Attorney's closing line. Between the trial and the sentencing, two further cases of gun/taser confusion were brought forward where the police officer was actually carrying the taser opposite to his gun. One of them occurred in Kentucky one year before the incident in Oakland, involving an officer with twice as much experience. According to the Court, the "unexpected" findings did not merit a new trial but they did bolster the defense's argument.

Catastrophic slips often look like intended behaviors, a manifestation of the agent's ill will. This is both a function of the "obviousness" of the mistakes, the harms they cause and, in some cases, the availability of potentially explanatory narratives. To point out the obvious, the case under discussion took place in the context of a well-documented history of police abuse and racial discrimination. So, many interpret Mehserle's actions in that light. Once these details come in, however, things often look more complicated.

Slips, as we shall see, are mistakes even from the point of view of those who make them. They are not, as Mehserle's defense would have it, mere accidents that happen to us. If they were it would be hard to make sense of the fact that people tend to be held responsible for them.<sup>4</sup> Instead, they are instances of agency—specifically, instances of intentional agency. And whereas they do not necessarily impugn the intentions of the agent or her regard for the consequences of her actions, they are not simply unlucky events. To the contrary, as it will become apparent below, they often result from a long history of *explicitly sanctioned* behaviors. So, even if they are instances of individual agency, they speak too about the role that institutions can have in shaping what we do.

Before proceeding, let me briefly introduce a methodological comment. Evidently, slips do not fit easily into the categories defined in our standard philosophical theories. But, as philosophers, I believe this is something that we *ought* to address. Mehserle's slip is not a sanitized thought experiment concocted to rebut a piece of orthodox theory but a real-life

---

<sup>4</sup> Calling something a "slip," therefore, is not to excuse it or even to diminish the gravity of its occurrence. However, discussing responsibility (moral or legal) for slips is beyond the scope of this paper. See Amaya & Doris (2014) for discussion.

mistake that needs to be explained and, more importantly, that needs to be prevented from happening again. Hence, we *must* be able to say something about it that goes beyond the obvious platitudes. Clearly, in the absence of a detailed understanding of what kind of mistake it is and why it happened, it will be hard to devise effective ways of preventing further occurrences of it.

## Slips

Let us begin then with a preliminary characterization of slips.

Slips are common mistakes. You set out to do something that you normally know how to do, and yet, without changing your mind, you wind up doing something else. Fortunately, most slips are inconsequential; some are even endearing. My father, for instance, used to call me by the name of his youngest sibling. While they were growing up, looking after him was my father's responsibility.

In comparison, episodes of police officers shooting unarmed suspects due to gun/taser confusions are uncommon.<sup>5</sup> Still, just looking at the institutional responses to these shootings, as opposed to deliberate ones, generate is enough to give one pause. Whereas Mehserle was ultimately convicted of manslaughter and sentenced to two years in prison, Darren Wilson, who intentionally shot Michael Brown several times in Ferguson, Missouri, was exonerated twice of criminal wrongdoing.

Although slips come under different guises, they occur in discernible patterns (Norman 1981; Reason 1984; Amaya 2013). Every slip requires having an intention in mind, say, the intention to write a check to pay a bill, or to subdue a suspect by tasing him. Yet, what you do then is inappropriate in the light of that intention and the information you possess. You inadvertently date the check with the wrong year. Or you subdue the suspect by drawing the weapon from the wrong holster. Given the information you have, you wind up acting contrary to the intention behind your action.

---

<sup>5</sup> Between 2001-2009, at least nine cases of gun/taser confusions were reported in the US, according to AELE Law library. See "Weapon Confusion and Civil Liability." 2012 (6) AELE Mo. L. J. 101.

We can make this more precise if we think of slips as performance mistakes (Amaya & Doris 2014). In them, the error does not lie with your judgment or your willingness to stick to it but rather it lies with the way you carry your intention out. Thus, as a Chicago police officer did in April of 2009, you could decide to tase someone, *judging wrongly* that he is attacking you—he might be simply having a diabetic seizure. Or, as it occurred in Honolulu in September of 2010, you could get angry at a suspect, *fail to exercise self-restraint*, and tase him for 30 straight seconds—the rule is not to exceed 5 second-bursts. These actions are surely mistakes. Neither, however, is a slip.

Obviously, different mistakes can pile up on occasion. As the prosecution argued, it is likely that the attempt to arrest Grant was “cheap”: apart from a technicality, there was no probable cause. Some witnesses claimed that, after being warned of the arrest, he became cooperative. So, under those circumstances, even using a *non-lethal* weapon against him was unlawful. Mehserle and his partner, it seems, escalated a situation that could have easily been handled differently. Given how widespread racial prejudices are, these are live hypotheses.<sup>6</sup> Still, having made the decision to arrest and tase Grant, and having how pinned down on the ground, Mehserle made a further mistake in following a well-internalized routine. In doing so, he slipped.

Now, unlike other performance mistakes, slips do not involve a mere failure to implement an intention. They involve an *incorrect* implementation of an intention, which is something different. In the slip, you do not simply let the opportunity to act as intended go by; the intention actually prompts your behavior. You normally have the necessary skills to execute it correctly. And, typically, there are no external factors preventing you from doing it. The mistake, therefore, is not excusable by forgetfulness, lack of skill, or extreme hardship.

Consider how Mehserle reached for his weapon. It was a display of well-coordinated, and skillful action. He warned Grant that he was going to tase him, he pinned him down,

---

<sup>6</sup> In cases like this where there is limited amount of information, it is hard to know how influential these prejudices really were at the individual level (Payne et al. 2018). Even so, accepting what has been said so far, the presence of racial biases in Meshserle would help explain why he made the decision to subdue Grant but would not explain why, having made that decision, he slipped and grabbed the wrong weapon.

shouted at his partner to clear the way, and pulled the trigger once. It was not as though he was forced to do any of this. He not did act fumblingly or in a physically awkward way. If Mehserle had intended to shoot Grant, he would have probably done something similar to what he did.

### **Doxastic guidance**

Accounts of intentional agency often assign beliefs an essential role in the production of action. The approach can be traced back to Aristotle's claim that actions are the conclusions of practical syllogisms. But it is also common among contemporary action theorists. Roughly put, the idea is that if someone acts with an intention, what the person does at the time is guided by her beliefs about how to achieve what she intends. The beliefs guide the action in the sense that they identify the means for achieving what is intended with it, explaining and rationalizing her behavior. For convenience, I shall call this *the principle of doxastic guidance*.

Theorists who seek to reduce intentions to desires and beliefs have endorsed versions of the principle. Briefly, if intentions are combinations of desires and beliefs about how to satisfy them, acting with an intention seemingly requires acting in accordance with those beliefs. Still, one need not be a reductivist about intentions to agree with this. One might hold that intentions are *sui generis* states that guide action, but insist that they play this role by comprising plans that reflect the agent's instrumental beliefs.<sup>7</sup>

Before getting into the discussion, we should get clear on what the principle of doxastic guidance claims. First, the principle concerns actions rationalized by personal-level attitudes and states. Second, it concerns actions that have a teleological structure, in the sense that they are done for the sake of some end. It is, in other words, a claim about *instrumental rationalization*. The principle is meant to explain how, when an agent acts intending to achieve some result, what she does can be seen as instrumentally rational, at

---

<sup>7</sup> Bishop (1989) and Enç (2003) provide examples of non-reductivist positions about intentions committed to a version of doxastic guidance. Mele & Moser (1994) are an exception: following Brand (1984) they claim that action plans need not be doxastically structured.

least from the position occupied by her. According to it, what plays this rationalizing role, connecting actions and intended ends, are the agent's instrumental beliefs.

There would seem to be various counter-examples to the principle. But given the clarification above, it is easy to see why they are merely apparent counter-examples. Consider, for instance, *arational* actions (Hursthouse 1991; Betzler 2009). Although these actions are not rationalized by instrumental beliefs, they do not seem to have a teleological structure. Hence, they are not counter-examples to doxastic guidance. Another case is *skilled motor actions*. To the extent that motor schemata control the fine details of skilled movements, beliefs might only indirectly guide them (Clarke 2011; Brownstein & Michaelson 2016). Being comprised of sub-personal representations, however, motor schemata are not suitable candidates for rationalization.

Slips, by contrast, are exceptions to the principle. There you act with an intention to achieve some end, but your behavior fails to align with your beliefs about how that could be done. The behavior, in other words, has a teleological structure, given by personal-level attitudes and states, which rationalize it. But that structure cannot be reduced to what you intend and believe then.<sup>8</sup> As Mehserle's case illustrates, in these cases what forges the connection between intention and action are *not* one's beliefs but one's acquired habits.

It would certainly beg the question against doxastic guidance if the present challenge were to depend on an idiosyncratic account of belief. But it doesn't. Here, beliefs need not even be identified with overt judgments, occurrent thoughts, or propositionally structured states. As we shall see, the patterns in which slips occur create problems for the relevant belief attribution, even granting a broad *dispositional* understanding of belief. Likewise, it would be unfair if the challenge presupposed a radical departure from standard ways of thinking about intentional agency, for instance, by severing its connection with acting for reasons. But it doesn't. Habits, as understood here, are dispositions to execute intentions in

---

<sup>8</sup> Obviously, even when a person slips, some beliefs will guide the execution of her action. The point is that the guidance is not exhausted by the agent's beliefs; habits play a role too. That is why, as we shall discuss below, the slip cannot be reduced to a doxastic mistake.

ways which have worked in the past and that have been internalized through a process of rehearsal.<sup>9</sup> Acting habitually in this sense, as we shall see, is one way of acting for reasons.

Advocates of the principle might respond to the challenge in, at least, two ways. First, they might claim that the slip is not a case of intentional agency. That is, they might accept that the person doesn't act guided by her beliefs but deny, for various reasons, that the resulting behavior is one that qualifies for rationalization. Alternatively, they might argue that, even though the action isn't in line with what the person would accept upon reflection, it is in line with (and, hence, rationalized by) her beliefs. Her beliefs, at least those that she had at the moment in which she slipped, just so happened not to be aligned with her more reflective attitudes.

In what follows, I discuss *prima facie* plausible ways of developing these responses, showing why they are not as promising as they might initially seem. With respect to first kind of response, I argue that slips have an intentional description, that they are not instances of deviance or habitual interference, and that there is a robust sense in which they count as episodes of acting for reasons. With respect to the second, I show why the reduction of slips to doxastic mistakes is unworkable, once we take seriously the differences among the dispositions that make up habits and beliefs.

### **Under a description**

We begin with the first kind of response.

According to Donald Davidson's famous *dictum*, if a person acts with an intention, the action can be described in a way in which it is intentional (1980: 46). Consider, to use his example, the navy officer who sinks the Bismarck by mistake, falsely believing that the Tirpitz (not the Bismark) is the approaching ship. Under the description "sinking the

---

<sup>9</sup> Thus understood, habitual behaviors need not be observed with the frequency with which, say, some compulsive behaviors (e.g. biting one's nails) can be seen. The frequency in which they are manifested, as it will become clear below, depends upon how often the person forms the type of intention the habit is supposed to subserve. See Douskos (2019: 4312-4313), who uses for these purposes the term "habitual routines."

Bismarck” his action is *not* intentional. It is intentional, though, if it is described as sinking the approaching ship.

Some philosophers have argued that no intentional description of the slip can be given (Peabody 2005; Pollard 2006). Mehserle had the intention of subduing Grant. He deployed his gun by mistake. But, to the extent that he did not deploy it guided by false beliefs about it (say, where the taser was, etc.), it would seem that one cannot describe the mistake to make it come out as intentional. Thus, assuming the truth of Davidson’s *dictum*, we would not have a case of an agent acting with an intention.

It’s easy to see where these philosophers are coming from. Unintentional descriptions of the slip are easy to come by. For example, Mehserle unintentionally shot Grant. Some ways of describing what he did are intentional, but are not descriptions of the mistake as such, say, Mehserle pinned down Grant and shouted at his partner to clear the way. Other descriptions are simply too general. Hence, it is not obvious that they make reference to the mistake and, if they do, whether they are intentional descriptions of it: Mehserle assaulted Grant, he subdued him, etc.

It is possible, however, to find an intentional description by looking at how the actions in the slip unfolded. That is, instead of trying to describe the mistake in one sweep, one can focus on the actions that constitute it and describe it in a *piecemeal* fashion. To wit, Mehserle drew the weapon from his right holster, held it with both hands, and pulled the trigger. In doing these things he evidently made a mistake. Had he drawn the weapon from his *left* holster, say, or had he *clubbed* Grant with it, there would have not been a shooting. It would seem, nevertheless, that he did each of them intentionally.

Notice, first, that these were not things that *merely happened* to him. The gun did not drop from his holster into his hands, or go off on its own. Also, they were not activities *passively* or *idly* undertaken. Mehserle did not draw the weapon and pull the trigger as one absently taps one’s foot while working in the computer (O’Shaughnessy 1980; Steward 2009). Nor did he do these things *reflexively* as one extends one’s arms to break a fall (Brand 1984; Bratman 1987). As evidenced by his verbal warnings, he did them having an intention in mind.

Clearly, the execution of that intention was not thought through.<sup>10</sup> Perhaps, at some point Mehserle pondered *whether* to deploy his taser. But, likely, he did not pause to think about *how* to deploy it. He grabbed the weapon in his right holster but didn't ask himself *which* weapon was this. Instead, his actions were part of a routine that, as a police officer, he had internalized as part of his regular training. They were, in this regard, like the routine of pressing the clutch and shifting gears while driving. It was a script that allowed him to act without having to think about the details of his actions.

Let me qualify this. To the extent that Mehserle didn't shoot Grant on purpose, it would be a stretch to insist that he acted with the belief then that he had a *gun* in his hand. Also, as we shall discuss in detail below, it would be problematic to say that he acted as he did because he believed he had the taser in his hand. Other than allegiance to doctrinal guidance, attribution of this or similar beliefs seems gratuitous. It also creates problems interpreting Mehserle's overall behavior.

Yet, it is plausible to think that, under some description, Mehserle was aware of what he was doing. For one, he didn't unconsciously follow the routine of drawing the weapon from his right holster and pulling the trigger, say, as one pulls the covers to one's side of the bed in the middle of the night.<sup>11</sup> For another, he didn't draw the weapon from his right holster thinking that he was reaching for his *pocket*, or pull the trigger thinking that it was the *safety*. In fact, if any of these things had crossed his mind, given his intention to tase Grant, he would have paused to change course.

It is precisely because of this that we get the desired description. Mehserle, we can agree, might have not known that he was about to shoot Grant. But he knew enough to be able to realize what he was about to do. Importantly, Mehserle was not merely aware that

---

<sup>10</sup> See Douskos (2019), who distinguishes habits and skills, in these terms. In brief, the person who displays a skill is attentive to the way she is acting: *how* she is doing what she intends. The person who acts habitually is relatively unencumbered by the *how* of what she is doing.

<sup>11</sup> Although slips typically involve reduced attention to performance, this reduction does not result in some kind of unconsciousness antagonistic to intentional agency. See Bermúdez (2017), Christensen et al (2016), and Fridland (2014) for a discussion of how much consciousness there generally is in skilled and habitual behaviors.

he was following this routine, as an spectator would have been, seeing things from a distance. He intended to subdue Grant and because of this he followed a routine he had internalized as part of his training. It was this, not simply a confusion on his part, that led to the shooting.

Slips, then, are not just accidents; they are mistakes agents make. They are not passive doings; they occur because of the agent's active intent. And they do not happen because the agent is unaware of the details of his doings—although they could have been avoided had the agent reflected more on what she knew. They are, instead, cases where one forms an intention and acts out of a habit or routine as a way of executing it. By acting habitually, one winds up making an error. Yet, to the extent that one acts habitually in the pursuit of the intention, one also winds up *intentionally* doing various things.

### **Deviance**

For advocates of doxastic guidance the present considerations might not yet place slips squarely in the domain of agency. Perhaps they agree that slipping involves having an intention and acting habitually in response to it. However, for them, the fact that the agent didn't act in line with her beliefs could be a sign that the "actions" were not brought about *in the right way*. That is, even though the intention moved the agent to act, it caused her behavior *deviantly*, in a way that prevents it from being intentional, even under the proposed descriptions.

Causal deviance comes naturally in various forms. But, presumably, what matters here is what John Bishop (1989) calls "basic deviance." There, the agent forms an intention, acts in response to it, but the action is not shaped by the *content* of the intention (what the intention is about). It is merely a causal response to it. Davidson's mountain climber is the paradigmatic example: having formed the intention to let his falling partner go makes him so nervous that he accidentally loosens his grip and lets his partner go (1980: 79).

Slips are evidently different in some respects. In textbook cases of deviance, for instance, the behavior fits the agent's intentions, whereas the lack of fit is what is most striking about the present case. Also, slips need not result from an overtaking emotion of the kind that standardly figures in examples of deviance. Even if the scene at the platform

startled Mehserle, as an experienced police officer it is unlikely that he was too nervous to act as he should have.

We can, however, put these differences aside to properly address the objection. For, ultimately, the point of the comparison is not the agent's emotional state or the ultimate success of her doings, but their *problematic etiology*. In brief, in both cases the behavior depends upon the intention: it is caused by it via a certain psychological process. But the fact that those processes result in those specific behaviors is *accidental* with respect to the things intended in each case. What the agent does, in this sense, is caused by the intention but not really shaped by its content.<sup>12</sup>

Upon inspection, however, the similarity turns out to be only superficial. For, in the slip the content of the intention actually shapes how it gets executed, even if it doesn't fully determine the execution. First, even though some of the actions in it fall short of being adequate, others are simply *correct*. Second, the habitual routines behind the mistake are, as a rule, *semantically related* to their guiding intentions. Over a wide range of possibilities, had the intention differed in content, the behavior would have differed accordingly.

Various features of Mehserle's case are telling in this regard. As the Judge explained in delivering the sentence, many of things he did were in accordance with his intention to subdue Grant with his taser. He moved back to gain some distance from him, which is consistent with the fact that tasers, unlike firearms, do not work at extremely close range. Also, he pulled the trigger only once. He did not "double" or "triple tap," as police officers are trained to do when deploying their firearms.

Importantly, Mehserle did not follow any random routine. He did not, say, reach into his back pocket for his wallet or his car keys, as perhaps it was his habit to do every night after getting home. He did not try to subdue Grant by biting him, which might have worked as way of achieving the intended result. The reason seems clear: wallets and car keys do not

---

<sup>12</sup> There are various ways of diagnosing what is deviant about deviant chains. Here, I consider the diagnosis offered by Bishop (1989) in terms of content sensitivity, which is standard (certainly, not unanimous) among action theorists. Note, at any rate, that the goal here is not to advance a solution to the problem of deviance, but to show why assimilating slips to these cases will not help address the present challenge.

help subduing people; biting is not standard police procedure. Deploying the weapon next to one's dominant hand is, in contrast, a sanctioned procedure for subduing suspects.<sup>13</sup>

To see the significance of these points, let us distinguish two ways in which one's behavior might be out of line with respect to one's intentions and beliefs. One of them is shared with cases of deviance. There, the behavior is *not* sensitive to the content of the attitudes actively entertained by the agent. It is caused but not necessarily shaped by them. The other is characteristic of *cognitively under-specified* processes. Here, the behavior is shaped by the agent's attitudes but, because not all the relevant attitudes are actively entertained at the time, it winds up being under-specified relative to some of them. Hence, what the agent does reflects some of the relevant contents of her mind, but not all the relevant ones (Reason 1992; Sellen & Norman 1992).

Slips are cases of under-specification. This is why intention and action fail to be in line, given the agent's overall beliefs. It is also why the mistake semantically approximates what would have been appropriate behavior. In brief, the agent sets out to act on an intention that can be executed in a number of ways. Normally, she has enough information to execute it correctly. But at the time only a subset of that information is present in her mind. As a consequence, the non-specificity of the intention is not correctly resolved: it is resolved adequately enough to rule in various related implementations, but not enough to rule out all the inappropriate ones.

Think about our case again. In principle, a police officer carrying a taser has various ways of subduing a suspect. Which is the correct way of doing it depends upon several things: where the weapon is, how to deploy it, etc. Obviously, at the time of the shooting, Mehserle knew all this. The problem was not lack of information, but rather that the information he had didn't become active in time to prevent him from following an internalized routine. He was trying to act quickly and many things were happening at the same time: he was trying to cuff Grant, his partner was shouting aggressively, people on the platform were protesting the arrest, etc.

---

<sup>13</sup> Diary studies suggest that this is a robust generalization. In the slip, behavior deviates in familiar, context-relevant, often socially sanctioned ways (Reason & Mycielska 1982; Jónsdóttir et al. 2007).

This is not to say that the mistake was bound to happen, or that Mehserle was a mere victim to the situation. Not only, as mentioned earlier, it possible for slips to occur among other mistakes but sometimes they are actually *precipitated* by other mistakes. And there is good reason to think that this was the case here. In brief, by the time a group of police officers gets in the position in which Mehserle was then, it is reasonable to conjecture that lot of mistakes had already been made.

### **Interference**

As we've seen, slips are not instances of causal deviance. But, some philosophers might argue that their occurrence suggests a broadly similar diagnosis. In deviant cases, some event beyond the person's control (say, her nervousness) *interferes* with the way the she was disposed to execute her intentions (Enç 2004; Aguilar 2012). Similarly, in the slip, the habit interferes in with agent's the exercise of her agency. As a consequence, what she does then doesn't really count as falling under the umbrella of the doxastic guidance.

It is easy to see the force of the objection if we focus on some behaviors that resemble slips and that clearly fit this description. With rehearsal, a familiar routine becomes automatized; once automatic, it operates *autonomously* of the agent's control, even to the point of interfering with it. The classic Stroop task provides an illustration. Although subjects are asked to report the color of a word visually displayed, they easily wind up reporting the color named by the word (Stroop 1935; MacLeod 1991). There is a widely accepted explanation for this. Among literate people, reading is so internalized that the mere presence of words is enough to override the attempt to report the observed color.

Slips embody some form of automaticity, which explains why some theorists have likened them to performances in the Stroop task (Norman 1981; Sripada 2019).<sup>14</sup> While they happen one normally doesn't deliberate about or really choose how to act, one typically acts unencumbered with the details of one's performance, etc. At the same time, the

---

<sup>14</sup> Here I follow the usual practice of identifying automaticity by the concurrent presence of certain features, none of which is by itself necessary for automaticity (Stanovich 2004; Moors & De Houwer 2007).

automaticity of the slip differs significantly from the one exemplified in the Stroop task, which admittedly interferes with intentional control. In brief, rather than being autonomous from the agent's intentions, the routines and habits in the slip are, in fact, subservient of them.

To see this, we should notice several things. First, unlike the situation in the Stroop, the slip does not involve a competing *concurrent* attempt. Even if the agent would have done otherwise, at the time there are normally no signs of her *trying* to do it. Judged by the deftness of his movements, for instance, it does not seem that Mehserle was trying to reach for his *left* holster when the shooting occurred. If anything, to the extent that his actions unfolded effortlessly, the opposite seems to be true.

It is true that in the slip the agent acts in ways that bypass what she otherwise would have wanted to do. But this does not make the habitual performance something beyond her control. For one thing, the habit is not mandatory or even hard to override. In fact, one of the hallmarks of the slip is that it has a *quick and easy cure*. Typically, awareness of an imminent slip is enough to prevent the mistake (Baars 1992; Amaya 2013). Performance in the Stroop task differs sharply. Even subjects cognizant of the task cannot avoid making mistakes or, at the very least, slowing their reading times.

Actually, there is an important sense in which acting habitually is under the control of one's intentions, a sense in which reading a word in the Stroop task is not. To be read, the word just needs to be in your visual field; the intention to comply with the instructions of the task just facilitates the *stimulus-driven* response. In the slip, by contrast, the intention is *inductively connected* with the behavior via a successful history of practices, a history that explains why the habit was acquired in the first place. Simply put, the behavior became habitual because it was an acceptable, or at least a sanctioned, way of executing intentions of the kind that prompted the agent to act.

Consider Mehserle again. As a right-handed police officer, reaching for the weapon on your right holster can be an appropriate routine to have, even a life-saving one, when subduing an aggressive suspect. At the very least, this is why he was trained to internalize it, given hundreds of opportunities to practice it. Once you reach the decision to deploy your weapon, you do not have to think too much. As the defense expert brought for the trial

put it, “muscle memory” takes over in ways that allow officers to react quickly and without hesitation.

Without doubt, many considerations speak against internalizing this routine. Arguably, using a lethal weapon should be a last resource when subduing a suspect. It should not be something that you do without having to think too much, much less without proper assurance that the suspect is really being aggressive towards you (not simply startled by your response to him). Yet, with the introduction of tasers as part of police equipment, the routine became a potential source of trouble for entirely different reasons. Here, as in other cases, changes in the world might *undercut* inductive support. With two weapons in your holster, you’d better not follow automatically the routine you acquired when you were carrying only one of them.

We should not underestimate this point. I’ve been arguing that Mehserle’s slip should not be considered a mere accident but an exercise of his agency. This, however, should not make us blind to the fact that exercises of our agency are shaped not just by our internal psychology (whether this is made up by beliefs or by habits) but also by the institutional arrangements and practices surrounding what we do. In fact, as we will now see, it is precisely in the light of these arrangements and practices that our habitual actions often come to be, even if they turn out to be glaring mistakes, in some respect reasonable.

### **Reasonableness**

As mentioned at the outset, the idea of doxastic guidance is meant to apply to actions *done for a reason*. At this point, it might be questioned to what extent the actions in the slip fit this description. To be sure, being an inadvertent mistake, the agent does not act having in mind a reason for doing otherwise. Her behavior is *not irrational* in that sense. But, in the absence of a relevant belief connecting her actions and her goals, one could object, there is nothing functioning as *the agent’s* reason for doing what she did.<sup>15</sup> Hence, even if slips are instances of acting with an intention, they do not really constitute a challenge to doxastic guidance.

---

<sup>15</sup> For the contrast between acting out of habit and acting for a reason, see Ryle (1949: 132), Goldman (1970: 91), Pollard (2006), Arpaly & Schroeder (2014: 80-85), and Chan (2016).

Obviously, some habitual behaviors are *not* done for a reason.<sup>16</sup> There are also no considerations favoring them; if they are, they do not explain why the agent acted as she did. Yet in the kind of case considered here the agent acts motivated for a reason: the goal she intends to achieve. And, to the extent that the habit was internalized in the pursuit of intentions of that kind, the agent also has some reason (a bad reason, but some reason nonetheless) to act in accordance to it. Acting habitually, at least in such cases, is recommended by what person currently intends and by an internalized history of *successful or sanctioned* practices.

Let me explain this last point by contrasting two ways of acting for a reason. According to a first way, an agent acts for a reason only if she judges, or would judge, her action as justified by the relevant considerations at hand, that is, by what she intends and believes then. Whether the judgment is understood as actually taking place or merely counterfactually, it is hard to see the slip as an instance of acting for a reason in *this* sense. Arguably, Mehserle did not have time to ask himself whether deploying the weapon in his right holster was the correct thing to do. Likely, based on what he believed, had he asked himself the question he would have immediately realized his mistake.

According to another way of seeing the matter, however, an agent acts for a reason whenever her action is the product of a *reasonable disposition*. Briefly put, a disposition is reasonable if it generally tracks reasons and tracking those reasons explains, in part, why it was acquired. It is in this second sense that acting out of habit, even if it results in a slip, can count as an instance of acting for a reason (which is not the same as saying that it is an instance of acting for a good reasons, much less of doing what one has most reason to do). Whereas in some circumstances following it might lead one astray, if the habit is inductively supported by a history of *successes*, there are general considerations speaking in favor of acting out of it. In so far as that history explains why it was acquired, those considerations can function as the agent's reasons.

We can put this in connection to what was said earlier. As argued above, slips occur due to cognitive under-specification. Under-specification can have various sources: say,

---

<sup>16</sup> Habituation might be the product of a variety of interventions (e.g., classical conditioning, pharmacological inductions, etc.) that lie “outside the space of reasons.” It is in these cases like this that acting out of habit is clearly not acting for a reason.

one might be psychologically unable to retrieve information one has. In the case of slips, however, it is a function of the fact that bringing relevant information to mind is a *resource intensive* exercise. It is easily impaired if at the moment of the action one's mind is elsewhere or, as in Mehserle's case, if one is forced to act under significant time pressure.

It is in light of these limitations that habits can connect an agent to her reasons. By *standing in for* beliefs in resource demanding situations, they dispose the agent towards ways of executing her intentions that have worked or have been sanctioned in the past. Overall, the results tend to be good: provided that the circumstances of internalization are congruent enough with the contexts of action. The downside, of course, is that if the circumstances are not congruent, and the agent overlooks the difference, a slip can result.

Slips, then, do contrast with some paradigmatic instances of rational agency. But the contrast is not between acting for a reason and acting habitually. To the contrary, it is between two ways of understanding what it is to act for a reason—or, more precisely, two ways in which an agent might be in contact with her reasons.<sup>17</sup> One of them, which goes hand in hand with the principle of doxastic guidance, focuses on the *individual's present attitudes*. The action is done for a reason because it is guided by a portion of the agent's psychology that latches onto considerations that justify, by her lights, its performance.

The other notion of acting for a reason emphasizes not her present attitudes, but her *history as an agent*. There too the agent's current psychology, as embodied in her habits, latches on to considerations that justify her actions, given her intentions. Yet, what makes these considerations *reasons for her* is not a set attitudes presently held: her actual judgments or the beliefs that would support her counterfactual judgments. It is, instead, the fact that the habit was internalized because, given the agent's circumstances at the time, acting in those ways in the pursuit of those intentions was, *institutionally speaking*, an appropriate thing to do.

---

<sup>17</sup> Arruda and Povinelli (2018) note how the literature on reasons for action tends to focus on approaches based on endorsement and justification. The present proposal is an alternative to this standard approach, in line with what they call a “directing,” as opposed to an “endorsement,” relation. Note, at any rate, that not all cases of acting for reasons in the absence of endorsement are historical in the sense noted above.

Obviously, to say that there were reasons for Mehslere to act as he did is *not*, in any shape or form, to condone or excuse what he did. A behavior can be reasonable in the light of a person's beliefs, but the beliefs of the person might be false, perverse, or simply unreasonable. Likewise, in light of some institutional arrangement, a certain habitual behavior might be reasonable. But the existence of the institutional arrangement and the fact that it sanctions some behavior as within procedure is no excuse, morally or otherwise, for behaving poorly.

### **Dispositions**

I now turn to the second way of fending off the challenge we identified at beginning of our discussion.

According to it, even though slips are instances of intentional agency, they are *not* counter-examples to doxastic guidance. Rather, being used to follow a given routine, the agent comes to believe, on inductive grounds, that the behavior she is about to undertake (under the appropriate descriptions) helps achieve what she intends then. On this view, slipping is both acting with an intention and acting for a reason. But that's only because it is one way of acting under a false belief.

It is easy to see why reducing slips to *doxastic mistakes* might be appealing here. Doing it allows harmonizing their existence with orthodox action theory. Further, given a dispositional view of beliefs, such as the one many of these theorists hold, the reduction would seem independently motivated.<sup>18</sup> If, as argued above, the habits behind slips are inductively supported dispositions, slipping due to a habit seems like a way of acting in accordance to one's beliefs.

A closer look, however, shows that the reduction is problematic. Simply, attributing the agent a false belief yields predictions that are not generally observed when it comes to slips. That is, whereas Mehserle's behavior at the instant of the shooting might be consistent, say, with him believing that his taser was in his right holster, the latter is inconsistent with his overall behavior and the things he could have done then. Thus,

---

<sup>18</sup> For dispositional accounts of belief, see Armstrong (1973), Stalnaker (1984), Schwitzgebel (2010) and Buckwalter et al (2015).

attributing the belief helps rationalize the mistake at the cost of getting wrong the agent's larger psychological profile.

To see why, begin with the idea that beliefs are *multi-track dispositions* (Stich 1978; Evans 1985; Weiskopf 2008). They dispose one to act in ways that systematically depend upon one's background attitudes. Despite wide disagreements about the nature of beliefs, I take it that most theorists nowadays would accept this as definitory of beliefs. Historically speaking, this is why the behaviorist approach to them never worked out. Because beliefs guide action by becoming inferentially integrated with other attitudes, there are no behavioral markers for them. Their manifestations normally include vastly different pieces of behavior.

Evidently, some idealization is going on here. Inferential integration takes time and effort. Some beliefs therefore will inevitably fall short of exhibiting this kind of systematicity: they will be too short-lived or too marginal to ever get integrated. Still, even within those limitations, beliefs tend to have a broad range of potential influence. At a minimum, because they can be extrapolated and combined with other attitudes, their potential influence extends *counterfactually* beyond a circumscribed type of behavior or situation.<sup>19</sup>

With this in mind we can return the discussion of our cases. The first thing to note here is that the dispositions behind slips exhibit the opposite tendency. Slips are characteristically island mistakes, which means that beyond the agent's actions at the time, there is normally no systematic confusion to be explained. This much became evident in Mehserle's case: that reaching for the weapon in his right holster and deploying it when intending to tase a suspect was an isolated incident. He did not repeatedly confused his taser with his gun or, less dramatically, was prone to confuse left and right. Perhaps, if that were the case, there would have been some grounds for attributing him a false belief.

Suppose, though, that contrary to what I've said Mehserle shot Grant because he came to believe momentarily that reaching for the weapon in his right holster was

---

<sup>19</sup> Schwitzgebel discusses examples of beliefs that are not fully integrated with the behavioral trajectories of their subjects: what he calls cases of in-between belief. For such cases, he seems to agree with the point made above: "we believe that P," he says, "if our actions and reactions *generally* reflect a P-ish take on the world" (2010: 541).

appropriate means to tase him. Then, even if that belief were only short lived, it could have resulted in many other actions. Given his knowledge that he was carrying each weapon in an opposite side of his belt, for example, he could have reached for his left holster if he had instead decided to shoot Grant. Or combined with his belief that the taser was set up in a cross-draw configuration, he could have reached for it with his left hand.

No doubt, there is much of Mehserle's psychology that nobody knows about. Still, these imagined scenarios are unlikely. For both of them would have him behave in ways that run contrary to a generalization empirically confirmed now and then. Namely, slips show robust *frequency-effects* (Reason 1990, ch. 3; Sellen 1990): the appropriate actions tend to be replaced not just by associated ones, but by those that have been more intensely rehearsed in the past. A novel behavior taking the place of an appropriate one is certainly an exception.

The point can be made with a useful rule of thumb: if a mistake shows enough integration with the agent's present attitudes, that speaks against it being slip. In part, this is a restriction to avoid false positives. But it also reflects the psychology behind the mistake. The agent's behavior is the result of becoming disposed to follow a certain routine when having a specific type of goal in mind. So, unless there is a further story of habituation, one will not observe the agent making a slip that involves the same behavior but in the pursuit of a different type of goal. Nor will one observe the slip eventuating in novel behavior.<sup>20</sup>

If anything, Mehserle's mistake looks like an episode of lack of integration. His actions resulted from a disposition formed through rehearsal, which never got integrated with his acquired beliefs about taser deployment, the location of each of his weapons, etc. In fact, compared to those beliefs, his previously acquired routine seemed to be systemically isolated from many of his relevant attitudes. Significantly, during taser training Mehserle and the rest of the police officers were not allowed to bring their guns with them. In the ten hours that the training lasted, none of them ever got to practice using the taser while carrying a firearm on their weapon belt.

---

<sup>20</sup> Discussing the "Moses illusion," Sorensen (2011) argues that some slips can eventually result in variegated behavior. This is indeed possible. If a slip is not corrected, the agent can become *disposed to form a belief* accordant with her behavior. See Audi (1994) for the distinction between dispositions to believe and beliefs.

## Fragmentation

But one might question whether this lack of integration actually raises a problem for the doxastic interpretation sketched earlier. After all, human minds can be somewhat fragmented (Lewis 1982; Stalnaker 1984; Egan 2008).<sup>21</sup> Because one can fail to see how one's beliefs connect with each other, neglect some of their implications, or temporarily forget things, one's beliefs sometimes do not form a coherent whole. Some of them, in fact, can wind up non-accidentally isolated from the bulk of one's attitudes.

Clearly, psychologically realistic notions of belief need not live up to our rational ideals. Hence, we should not reject off hand the possibility of someone's mind being fragmented to the point of one of her beliefs operating only in a narrow type of context. At the same time, believing something involves being disposed to *regard it as true* (Railton 1994; Velleman 2000). Which means that even agents with fragmented belief systems are sensitive to pressures of integration and coherence. Far from being permanent partitions, doxastic fragments tend to get integrated with the agent's overall psychology when the right sort of information is presented to her.

The slip, however, is also different in this regard. Surely, the mistake is evidence of the agent's mind being fragmented in *some sense*. At least, it indicates that some of her behavior and beliefs were not properly aligned. Yet, their misalignment is not simply the consequence of the way information circulates in one's mind, but of something deeper. Unlike beliefs, habits are *past-dependent dispositions*: they dispose agents to act in accordance with a history of internalization and rehearsal. And, because of this, they tend not be sensitive to the kind of evidential pressures that normally shape one's doxastic attitudes.

---

<sup>21</sup> Dispositionalists about belief have traditionally been the advocates the fragmentation hypothesis. There is a recent representationalist version of it in Mandelbaum (2014) and Quilty-Dunn & Mandelbaum (2017). The main issue dividing these approaches is whether beliefs are relations to structured representations, not so much the issue of fragmentation. Accordingly, their discussion can be set aside here.

Think about the things that normally work (or not) as measures to prevent the slip. These are not the things that typically serve as antidotes for fragmentation. For instance, unlike cases where a person fails to see how her beliefs can add up to an inconsistency, explaining the conflict is normally unnecessary to cure a slip. Once you attend to it, the misalignment of behavior and beliefs becomes obvious without explanation. Likewise, anticipating the consequences of one's actions or being reminded ahead of time about what counts as an appropriate implementation of one's intentions are hardly effective methods to prevent the slip. It is much better to *re-train* oneself or to set up physical barriers that prevent one from following the habitual routine.

This was a controversial aspect of Mehserle's trial. As the prosecution insisted, during their training period officers were explicitly warned about possible gun/taser confusions. To avoid them, they were instructed to carry their weapons on opposite sides of their holsters (as Mehserle was doing the night of the shooting) or to set their tasers up to be reached with their non-dominant hand. The assumption was that if the officers were warned and the two weapons were far enough apart they were not going to get confused about their deployment.

Unfortunately, none of this made a difference in the practical training undergone by the officers. As we have seen, they were not allowed to train while carrying their guns. Perhaps more surprising, anticipation of the mistake made little difference in actual police procedure. At the beginning of every shift, they were randomly assigned holster belts with either of two possible configurations: each weapon on a different side of their body vs. both weapons on the same side with tasers set up for non-dominant draw. This prevented the officers from internalizing new routines that would have kept apart, as much as possible, taser and gun deployment.

We can put the point in more precise terms. To the extent that beliefs are truth-oriented dispositions, they tend to be sensitive to global pressures of coherence. For the same reason, belief fixation and revision tend to be *isotropic* processes: any relevant information can potentially lead to acquiring or abandoning a given belief (Fodor 1983; Burnston & Cohen 2015). By contrast, being past-dependent, the dispositions behind slips tend to be impervious to coherence pressures arising from upcoming or newly acquired information. That is why, although becoming aware of an imminent slip might be enough

to block the manifestation of the habit, learning that a habitual routine is inappropriate does not make the disposition to follow it go away.

Again, we should be careful with the way we handle exceptions here. Some episodes of belief acquisition are apparently unencumbered by evidence (Gilbert 1991; Mandelbaum 2014). Entertaining a proposition, in those cases, automatically leads to believing it; rejecting it, by contrast, requires *effort*. Also as cognitive dissonance theorists have noted, acquired beliefs can sometimes be impervious to coherence pressures. Under certain circumstances, for instance, agents are known to *persevere* in their beliefs, despite receiving good evidence of their being incoherent (Ross et al. 1975; McFarland et al. 2007).

Interestingly, rather than supporting a doxastic interpretation, these exceptions provide an even sharper contrast. Agents who slip *effortlessly* recognize their mistakes once these are pointed to them. What seems automatic in their case is the *recognition that they knew better*. On the other hand, agents who persevere in their beliefs often try re-interpreting their behavior or coming up with auxiliary hypotheses to harmonize the conflicting beliefs. This, however, is the opposite of what is commonly observed in the slip.

Consider Mehserle one last time. As soon as he shot Grant, he brought his hands to his head. Although nobody knows what went through his mind then, it does not seem that he doubted he had made a mistake. Mehserle did not try to find an excuse for what he did (the gun went off by accident). He did not try to justify the use of his firearm (I felt I was in danger). Nor even was he left in the train platform wondering where things went wrong (where was my taser again?). It was as though he had no way of rationalizing his behavior. As one witness of the incident described him: “The officer who pulled the trigger was in shock. [He] had a look on his face similar to ‘Oh my god! I can’t believe this happened.’ Like the deer in the headlights look.”

.....

Let us take stock. I have argued so far that slips are instances of acting with an intention. While the actions that constitute them are guided by the agent’s habits, as we have now seen, they do not seem reducible to her beliefs. Habits, unlike beliefs, are not multi-track or truth-oriented dispositions. Therefore, they do not exhibit the patterns of integration and the openness to evidential pressures characteristic of beliefs. This, I think, is sufficient to

establish the major point that we initially set out to make about slips being exceptions to doxastic guidance. Before concluding, however, we should address two further objections that cast doubt over the significance of the present challenge. Discussing them will also allow us to generalize the points made so far with respect to exercises of agency that do not necessarily involve slips.

### **Rationalization**

Doxastic guidance, as mentioned at the outset, is a principle concerning instrumental rationalization. So, one way of resisting the force of the challenge compatible with what has been argued so far is to oppose up front the idea that habits, as discussed here, can play the rationalizing role uniquely assigns to beliefs. There are various ways of developing this line of thought. But Nomy Arpaly and Timothy Schroeder (2014: §3.7) have advanced a view of this sort worth considering here. According to them, unlike beliefs, habits do not have contents and, because of this, they cannot rationalize actions. Hence, although habitual actions might count as intentional and habits might not be reducible to beliefs, acting habitually is not a candidate for rationalization.

Obviously, the notion of content is a complicated one. It is also disputable what exactly it means for a belief, dispositionally understood, to be contentful. Still, even granting that habits do not have contents, the skeptical conclusion Arpaly and Schroeder push does *not* follow. It is certainly true that habits, by themselves, cannot rationalize actions. But, together with the intentions motivating the agent to act, they can rationalize what she does.

Think in more detail about the idea of rationalization. First, rationalization is supposed to mark *a logical relation*. This is why the practical syllogism often gets invoked in this context. As it is often put, intentions and beliefs are meant to relate to the intentional description of the action they rationalize in the way that the premises and conclusions of the syllogism are related to each other. In addition, rationalization is supposed to occur in *an explanatory context*. That is, whatever rationalizes the action does not merely show that the action makes logical sense; its being thus logical for the agent also has to be part of the explanation of the action.

As we have seen with Mehserle, agents sometimes cannot offer a rationalization for their slips. That is, they often find themselves unable to articulate a justification for what they did, or offer any sort of explanation that would make their behavior a bit less puzzling. This, however, should not be taken as evidence that there is no logical structure to their behavior, or that such structure does not play a role in producing it. An agent can act in way that seemed rational to her, even if she unable to articulate afterwards exactly why it seemed so to her. An explanatory rationalization, in other words, need not be available for first-person report.<sup>22</sup>

Start with the logical aspect of the rationalization relation. Let us grant that in the absence of beliefs, there can be no *deductively valid* pattern of reasoning by reference to which the relation between intentions and actions can be articulated. Plainly, if habits do not have contents, they do not to have associated truth conditions. Still, there is an acceptable pattern of reasoning connecting the action and the intention that serves the purposes of rationalization. Specifically, an agent in the habit of performing some type of action by way of executing a general intention can reasonably conclude *by default* that there is reason to perform that action on the grounds of currently having an intention of that kind.<sup>23</sup>

This type of reasoning need not happen in one's head. But the same holds with respect to actions rationalized by beliefs: agents need not go through the syllogism that rationalizes them. The point of invoking the argument is *not* to describe a psychological process but to make clear the logical relation holding between the items over which a psychological process ranges. The argument is meant to display the logic that led the agent to see the action as a reasonable way of executing the intention.

In the slip, then, habits are related to intentions and actions in the way that defaults relate premises and conclusions in a piece of default reasoning. That is, they do not

---

<sup>22</sup> This is not because the agent cannot access her state of mind at the time of the mistake. She can access it in whichever way it is required by first person knowledge. Still, having realized that she made a mistake, she might not be able articulate a coherent picture of her state of mind at the time. What was I thinking?

<sup>23</sup> For discussion of how default processes can shape action at different levels, see Bach (1984) and Pollock (2008), who follow Reiter's formal treatment (1980).

function as premises from which the conclusions can be derived, say, the premise that some available course of action is the best way to satisfy one's current intentions. Instead, by reasonably disposing the agent towards some ways of executing those intentions, they sanction a straight transition from intentions to actions. As in a piece of default reasoning, they set the actions as defaults to be followed when pursuing associated goals unless considerations for doing otherwise comes to mind.

Further, these transitions make the mistake intelligible: the agent errs because she follows a default way of executing the kind of intention she had in mind. They help see why a reasonable agent could have made the mistake: the considerations for doing otherwise didn't come to mind in time. Also, they also explain why it seemed desirable for the agent to act as she did. Acting on an acquired habit might not secure intention satisfaction but, very much like reasoning by default, it is good enough in demanding circumstances. Last but not least, focusing on these transitions can also help us see how the institutions that surround us can play a large role in structuring our agency. As Mehserle's case illustrates, they contribute to the rationalization of what we do by shaping up what we take to our defaults to be.

In sum, if rationalization is understood as something that only contentful states can do, then, as Arpaly and Schroeder claim, habits might be unable to play the rationalizing role traditionally assigned to beliefs. But, as we just have seen, the possibility of rationalizing an action need not be bound by this requirement. Some item in the agent's psychology can structure what she does by disposing her to make certain logical transitions from intention to action. When this is the case, as with the habits behind the slip, rationalization and, thus, action guidance need not involve beliefs.

### **Glitches**

The idea of doxastic guidance is part of a project of providing a *unified theory of rational agency*. If the preceding remarks are on the right track, however, unification of this sort seems rather unlikely. Sometimes beliefs rationalize our actions. But, other times, as when we slip due to a habit, what we do is rationalized by dispositions of a different kind. There, forms of reasoning different from the practical syllogism provide a model for understanding what rationalization involves.

Some action theorists will no doubt balk at this conclusion because of its implications for the unification project. Perhaps, slips are not oddities in the life of human agents: everyone has had first-hand experience with them. Yet, they are not the way things *ought* to be or the way things *normally* are. Here, these theorists would argue, lies the problem. One cannot derive an ought from an is. But neither can one derive from our mistakes a view of our rational agency. As far as successful exercise of our agency goes, doxastic guidance remains a suitable generalization.

It would take considerable space to give a complete answer to this objection. But, for present purposes, it helps to distinguish two things that are relevant to conceptualize a mistake. First, there are the standards by which the *performance* gets evaluated, that is, by which it can be said to result in a success or a mistake. These are standards by which, say, our performances are regarded as morally good, illegal, etc. There are, on the other hand, the standards by which one can evaluate *the processes* that lead to those performances. It is typically in terms of these that normal and abnormal processing gets distinguished, whether the abnormality is pathological or due to a temporary glitch.<sup>24</sup>

Considered in the abstract, it is clear that these standards can come apart. An unsuccessful performance may in principle result from an otherwise impeccable process. To take one concrete example from a different domain, in their landmark studies of inductive reasoning, Kahneman and Tversky (1973; 1982) observed how their subjects' intuitive judgments violated basic probabilistic principles. The violations, however, they claimed, did not seem to result from misapplications of the probability calculus, but rather from their reliance on a limited number of heuristics that would be reasonable to use in naturalistic scenarios.

In the case of the slip, the situation is analogous. What the agent does then is sometimes morally offensive or illegal; it is always contrary to her intentions and beliefs. Yet, it does not follow from this that the slip is the result of a glitch or a lapse in an

---

<sup>24</sup> The distinction between the evaluation of performances and processes lies at the heart of Herbert Simon's (1957, 1983) classic distinction between *substantive* and *procedural rationality*. Michael Bratman (1987: 5.2) also draws a similar distinction between the rationality of having general policies for (not) reconsidering plans and the adequacy of specific episodes of (non) reconsideration.

otherwise foolproof process of intention execution—the one that actually explains what happens in the successful case. There is, as we have seen, enough systematicity to the mistake to view it as the result of a process of intention execution of a *different kind*.

With this distinction in mind, we can now return the objection. There is, as the unificationist would claim, a sense in which slips are not the way things *ought to be*. The *performance* of the agent falls short of standards to which it makes sense to hold rational agents accountable, at least in principle: to execute one's intentions in line with one's beliefs. The objection, however, loses its force, once we focus on the *processes* by which our intentions can be executed. Slips, as we have seen, help dissociate two different kinds of psychological processes that are relevant here, each of which can support rational agency in a different way.

On one of them, beliefs direct agents to achieve their goals by disposing them to act in ways that make sense, given the information they have at the time. What they do then is rational in the light of that information. On the other, habits make salient certain courses of action as default ways of acting, on the grounds that those ways have worked in the past. In the light of our limitations and the constraints with which we act, this kind of inductive support speaks in favor of its rationality.

Importantly, even if most actions in everyday life accord with our beliefs, this is not enough grounds for thinking that doxastic guidance is what's normal there. Felicitous performances are poor at revealing overlooked distinctions at a process-level. Slips, on the other hand, by breaking the monopoly exercised by belief-based explanations of intentional action, open the door to *a flood of potential counter-examples* to doxastic guidance. Beliefs surely guide some successful performances. Yet, when beliefs and habits align with each other, for the reasons sketched here, one need not assume without additional evidence that the former, and not the latter, are the doing the guidance.

As I said, it would take much more space to develop these points properly. This would require, for instance, a more substantive discussion of the methodology of using mistakes in the re-construction of psychological process. It would also require spelling out a theory of everyday habitual successful agency. These are things that I cannot do here—I do them elsewhere. What has been said so far, however, should be enough to take the challenge slips raise seriously.

Obviously, in addition to the possibilities considered above, there are other ways of defending the principle of doxastic guidance. Say, one might qualify the principle with *ceteris paribus* clauses or one could try to explain the slip by attributing the agent some *tacit belief* that would guide her action. The problem with these responses is that, although they seem appealing at first, once developed, they lose their intuitive force.

It is not clear, for instance, what in general *ceteris paribus* clauses are supposed to exclude. However, if guidance by habits is not necessarily defective or abnormal, and it is possibly pervasive, as we have just seen, excluding it from an over-arching principle of instrumental rationalization seems ad hoc. Likewise, despite its ring of familiarity, it is hard to know what the expression “tacit belief” refers to, except something that is like a belief but falls short of being one.<sup>25</sup> It is harder to know how tacit beliefs rationalize and whether they do it in the way in which explicit beliefs are supposed to rationalize actions.

In fact, for the reasons just discussed, dismissing the challenge merely by gesturing at these possibilities seems needlessly conservative. By doing it one might get to preserve a cherished principle, or (at least) a version of it, or (at the very least) the letter of it. But, as we have just seen, one runs the risk of not fully appreciating the broad range of human agency. More importantly, one risks failing to see how easily avoidable and yet catastrophic mistakes, such as Mehserle’s slip, can be part of how we exercise our agency in the particular situations in which we find ourselves.

## Conclusion

Slips are exceptions to doxastic guidance. Or so I have argued here. I’ve shown that *prima facie* plausible ways of fending off the challenge do not work. Some of them seem plausible only because they overlook some features of slips that are necessary to explain what kind of mistake they are and why they happen. Others seem plausible only because they presuppose a narrow understanding of some core features characteristic of intentional agency.

---

<sup>25</sup> I can’t discuss here the semantics of *ceteris paribus* clauses or to develop an account of what a tacit attitude of belief (or desire) could be. But, for a sustained discussion of these issues in contexts similar to this one, see (Gauker 2005; Amaya 2013)

In the end, the discussion has served to sketch a model of how a psychology of habits can be integrated in an account of intentional agency. Acting habitually, as we have seen, is not antithetic to acting for a reason; habits sometimes function as rationalizing dispositions. Even though much remains to be said on this point, these are worthy results. With its reliance on doxastic guidance, traditional action theory has failed to recognize the distinctive psychology behind habitual action and with it the possibilities of erring that come with exercises of our agency.

Throughout this paper, I have discussed the 2009 killing of Oscar Grant by the police officer, Johannes Mehserle, and constructed an argument that has now shown that the incident was not merely an accident, but rather the product of a mistake the officer made. As deplorable as the officer's actions were, however, I have also emphasized that the mistake was not simply due to his individual state of mind, but also was, in part, due to a series of institutionally sanctioned practices, having to do with the way officers are equipped and trained. This is not to say that Mehserle had no control over his actions or could not have acted differently; obviously, much remains to be said regarding how our individual agency and our social arrangements operate in relation to each other. The points raised here, then, should be taken into account to evaluate not only what happened in Oakland that night, but they should also be employed to figure out what to do to prevent similar mistakes going forward.<sup>26</sup>

---

<sup>26</sup> Versions of this paper were presented at the Berlin School of Mind and Brain, the Gothenburg Workshop on Moral Responsibility, the Institute Jean Nicod, and the Universities of San Diego, Sheffield, Tübingen, and Washington University in St. Louis. Al Mele, Arnon Cahen, Sarah Robins, John Doris, Neil Levy, Roy Sorensen, Manuel Vargas, and Allison Wolff gave me feedback for which I am extremely grateful. I also want to thank the anonymous reviewers that helped shape this paper into its final version. Work on this paper was supported by grant #60845 from the John Templeton Foundation.

## References

- Aarts, H. & Dijksterhuis, A. (2000). Habit as Knowledge Structures: Automaticity in Goal-Directed Behavior. *Journal of Personality and Social Psychology*, 78(1), 53-63.
- Aguilar, J. (2012). Basic causal deviance, action repertoires, and reliability. *Philosophical Issues*, 22(1), 1-19.
- Amaya, S. (2013). Slips. *Noûs*, 47, 559-576.
- Amaya, S., & Doris, J. (2014). No Excuses: Performance Mistakes in Morality. In J. Clausen & N. Levy (Eds.), *Springer Handbook of Neuroethics* (pp. 253-272). Dordrecht: Springer.
- Armstrong, D.M. (1973). *Belief, Truth and Knowledge*. London, Cambridge University Press.
- Arpaly, N. & Schroder, T (2014). *In Praise of Desire*. New York: Oxford University Press.
- Arruda, C. & Povinelli, D. (2018) Two ways of relating to (and acting for) reasons. *Mind and Language*, 33(5), 441-459.
- Audi, R. (1994). Dispositional beliefs and dispositions to believe. *Noûs*, 28(4), 419-434.
- Baars, B. (1992). The many uses of error: Twelve steps to a unified framework. In B. Baars (Ed.), *Experimental slips and human error* (pp. 3-34). New York: Plenum Press.
- Bach, K. (1984). Default reasoning: Jumping to conclusions and knowing when to think twice. *Pacific Philosophical Quarterly*, 65, 37-58.
- Bermúdez, J. (2017) Do we reflect when we perform skillful actions? *Philosophical Psychology*, 30(7), 896-924.
- Betzler, M. (2009). Expressive Actions. *Inquiry*, 52(3), 272-292.
- Bishop, J. (1989). *Natural Agency: An Essay on the Causal Theory of Action*. New York, Cambridge University Press.
- Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. Cambridge, MA: MIT Press.
- Bratman, M. (1987). *Intentions, Plans and Practical Reasoning*. Cambridge: Harvard University Press.
- Brownstein, M. & Michaelson, E. (2016). Doing without believing: Intellectualism, knowledge-how, and belief-attribution. *Synthese*, 193(9), 2815-2836.
- Buckwalter, W., Rose, D., & Turri, J. (2015). Belief through Thick and Thin. *Noûs*, 49(4), 748-775.
- Burnston, D. & Cohen, J. (2015). In A. Raftopoulos and J. Zeimbekis (eds.), *The Cognitive Penetrability of Perception. New Philosophical Perspectives*. Oxford: Oxford University Press.
- Chan, D. (2016) *Action Reconceptualized*. Lanham MD: Lexington Books.
- Clarke, R. (2011). Skilled activity and the causal theory of action. *Philosophy and Phenomenological Research*, 80(3), 523-550.
- Christensen, W., Sutton, J. & McIlwail, D. (2016). Cognition in Skilled Action: meshed control and the varieties of skill experience. *Mind & Language*, 31, 37-66.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.
- Di Nucci, E. (2011). Automatic Actions: Challenging Causalism. *Rationality, Markets and Morals*, 2 (1),179-200.
- Douskos, C. (2019). The spontaneousness of skill and the spontaneousness of habit. *Synthese*, 196, 4305-4328.

- Egan, A. (2008). *Seeing and believing: perception, belief formation and the divided mind*. *Philosophical Studies*, 140(1), 47-63.
- Enç, B. (2003) *How We Act: Causes, reasons, and intentions*. Oxford: Oxford University Press.
- Enç, B. (2004). Causal Theories of Intentional Behavior and Wayward Causal Chains. *Behavior and Philosophy*, 32 (1), 149 -166.
- Evans, G. (1985). Semantic Theory and Tacit Knowledge. In *Collected papers* (pp. 322-342). Oxford: Clarendon Press.
- Fodor, J. (1983) *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fridland, E. (2014). They've lost control: reflections on skill. *Synthese*, 191, 2729-50.
- Gauker, C. (2005) The belief-desire law. *Facta Philosophica*, 7, 121-144.
- Gilbert, D. (1991). How mental systems believe. *American Psychologist*, 46(2), 552-585.
- Ginet, C. (1990) *On Action*. Cambridge: Cambridge University Press.
- Goldman, A. (1970). *A Theory of Human Action*. Englewood Cliffs, N.J.: Prentice-Hall.
- Hursthouse, R. (1991). Arational actions. *The Journal of Philosophy*, 88(2), 57-68.
- Jónsdóttir, M., Adólfssdóttir, S., Cortez, R. D., Gunnarsdóttir, M., & Gústafsdóttir, H. (2007). A diary study of action slips in healthy individuals. *The Clinical Neuropsychologist*, 21(6), 875-883.
- Korsgaard, C. 1997. The normativity of instrumental reason. In G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*. Oxford: Clarendon Press.
- Lewis, D. (1982). Logic for equivocators. *Noûs*, 16(3), 431-441.
- MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163-203.
- Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, 57(1), 55-96.
- McFarland, C., Cheam, A., & Buehler, R. (2007). The perseverance effect in the debriefing paradigm: replication and extension. *Journal of Experimental Psychology*, 43, 233-240.
- Mele, A., & Moser, P. (1994). Intentional action. *Noûs*, 28(1), 39-68.
- Moors, A., & De Houwer, J. (2007). What is automaticity? An analysis of its component features and their interrelations. In J. Bargh (Ed.), *Social psychology and the unconscious* (pp. 11-50). New York: Psychology Press.
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, 88, 1-15.
- O'Shaughnessy, B. (1980). *The Will: A Dual Aspect Theory*. New York: Cambridge University Press.
- Payne, K., Niemi, L, Doris, J. (2018) How to think about “implicit bias.” *Scientific American*, March 27, 2018.
- Peabody, K. (2005). Trying slips: Can Davidson and Hornsby account for mistakes and slips? *Philosophia*, 33, 173-216.
- Pollard, B. (2006). Explaining actions with habits. *American Philosophical Quarterly*, 43, 57-68
- Pollock, J. L. (2008). Defeasibly reasoning. In J. Adler and L. Rips (Eds.), *Reasoning. Studies of human inference and its foundations* (pp. 451-470). Cambridge: Cambridge University Press.
- Quilty-Dunn, J, Mandelbaum, E. (2018). Against dispositionalism: belief in cognitive science. *Philosophical Studies*, 175, 2553-2372.
- Railton, P. (1994). Truth, Reason, and the Regulation of Belief. *Philosophical Issues*, 5, 71-93.

- Reason, J. (1984). Lapses of attention in everyday life. In R. Parasuraman, & D. Davies (Eds.), *Varieties of Attention* (pp. 515-549). New York: Academic Press.
- Reason, J. (1992). Cognitive underspecification: Its variety and consequences. In B. Baars (Ed.), *Experimental slips and Human Error* (pp. 71-91). New York: Plenum Press.
- Reason, J., & Mycielska, K. (1982). Absent-minded? The psychology of mental lapses and everyday errors. Englewood Cliffs, N.J.: Prentice-Hall.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13, 81-132.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32, 880-8.
- Ruben, D.H. (2003). *Explaining Explanation*. New York: Routledge.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531-553.
- Sellen, A. J. (1990). *Mechanisms of human error and human error detection*. Thesis (Ph.D), University of California San Diego.
- Sellen, A., & Norman, D. (1992). The psychology of slips. In B. Baars (Ed.), *Experimental slips and human error* (pp. 317-339). New York: Plenum Press.
- Smith, M. (2010). The standard story of action: An exchange (1). In J. Aguilar, & A. Buckareff (Eds.), *New Perspectives on the Causal Theory of Action* (pp. 45-55). Cambridge, MA: MIT Press.
- Sorensen, R. (2011). What lies behind misspeaking? *American Philosophical Quarterly*, 48(4), 399.
- Sripada, C. (2019) The Fallibility Paradox. *Social Philosophy & Policy*, 36(1), 234-248.
- Stalnaker, R. (1984). *Inquiry*. MIT Press.
- Stanovich, K. (2004). *The Robot's Rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
- Steward, H. (2009). Sub-intentional actions and the over-mentalization of agency. In Constantine Sandis (Ed.), *New Essays on the Explanation of Action* (pp. 295-312). New York: Palgrave Macmillan.
- Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 45(4), 499-518.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Velleman, J.D. (2000). On the aim of belief. In *The Possibility of Practical Reason* (pp. 244-281). Oxford: Oxford University Press.
- Wallace, R. J (2001) Normativity, commitment and instrumental reason. *Philosopher's Imprint*, 1(3), <http://hdl.handle.net/2027/spo.3521354.0001.004>
- Weiskopf, D. (2008). Patrolling the mind's boundaries. *Erkenntnis*, 68, 265-276.
- Wood, D. & Neal, D.T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114(4), 843-863.