ON SELF-KNOWLEDGE OF MOTIVES

Pablo Hubacher Haerle
University of Cambridge, ph539@cam.ac.uk.

Abstract: Many philosophers claim that we have duty to know our motives. However, prominent theories of the mind suggest that we can't. Such scepticism about knowledge of one's motives is based on psychological evidence. I show that this evidence only mandates scepticism about knowledge of one's motives if we rely on a mistaken assumption which I call 'the myth of the one true motive'. If we reject this myth, we see that there is space to plausibly interpret the empirical data such that knowledge of one's motives is difficult, but not impossible.

> The so-called *motive*: another error.
>
> — Friedrich Nietzsche, *Twilight of the Idols*, §6.3.

We want to know who we are. Humans have a strong drive for self-knowledge (Strohminger *et al.* 2017). Moreover, numerous philosophers claim that we *should* know ourselves. For instance, Socrates held the Delphic imperative 'Know Thyself!' in highest esteems (Moore 2015: 7) and Kant thought that you ought to know "whether the source of your actions is pure or impure" (*MS* 6:441). Furthermore, Kant, and many philosophers after him, saw the duty for self-knowledge primarily as a duty to know one's *motives* (*G* 4:407; Ware 2009).

Yet, recent empirical evidence led many philosophers and psychologists to doubt whether we can know our the true causes behind our actions. For instance, a popular approach conceives of the mind as consisting of two systems: a conscious part—responsible for reasoning—and an unconscious part—leading us to act and think in spontaneous but advantageous ways (e.g., Wilson 2002; Doris 2015). Often, it is implied by such a view that what's really driving our actions lies outside 'the mind's eye' and can only be revealed by psychological experiments, neuronal modelling or evolutionary theorizing. What we ordinarily say and think about our motives is at best a lucky guess and at worst a misleading fiction (Mercier and Sperber 2011, 2017; Scaife 2014; Chater 2018).

Evidently, if true, such a view renders any duty to know one's motives problematic. After all, it is widely believed—including by Kant himself (*MS* 6:380)—that to have a duty to φ we must be *able* to φ. Does the duty for self-knowledge collapse, then, because it's impossible for us to know our motives?

I argue against this suggestion. First, I will briefly carve out what a motive is (§1) and what it means to know one (§2). Then, I'll present paradigmatic pieces of psychological evidence (§3) and show how they have been taken to suggest *scepticism* about knowledge of one's motives (§4). Next (§5), I argue that the psychological evidence only mandates such scepticism if we rely on a mistaken assumption which I dub 'the myth of the one true motive'. According to this myth, there is only one true cause of every action, such that if you fail to pick out that unique cause, you don't know your motive. Rejecting this myth allows for a reasonable interpretation of the empirical data, on which knowing your motives is hard, but not impossible (§6). As a result, the duty for self-knowledge—understood as an obligation to know one's motives—remains on a secure basis.

1. WHAT IS A MOTIVE?

Let's start with this passage by G.E.M. Anscombe:

> Consider the statement that one motive for my signing a petition was admiration for its promoter, X. Asked 'Why did you sign it?' I might well say 'Well, for one thing, X, who is promoting it, did …' and describe what he did in an admiring way. I might add 'Of course, I know that is not a ground for signing it, but I am sure it was one of the things that most influenced me' (1957: §13, p. 20).

If the signer is correct in believing admiration to be one of the things that influenced them most, it seems natural to say they have *knowledge* of their motives. How do we get such knowledge? And what kind of thing is a motive, anyway?

There are multiple ways to think about motives in philosophy (Anscombe 1957: §13, pp. 20f.; Velleman 1989: 199), psychology (Schultheiss and Brunstein 2010: 308: Thrash *et al.* 2010; Simler and Hanson 2018: 6ff.), and law (Heering 2015: 45f.; Simester and Sullivan 2019: 138). Some of these senses are very narrow—certain philosophers see motives as necessarily third-personal terms—while others are rather wide—some psychologists understand motives simply as tendencies to act in response to incentives. The legal understanding is a happy compromise. In law, whether an action was done with intent has important ramifications. It makes all the difference between murder and manslaughter. The *motive* behind an intentional action, however, is often seen as irrelevant to the law. An illegal action done with a good motive is illegal nonetheless (Heering 2015: 45f.; Simester and Sullivan 2019: 138). Behind this lies the view that motives are *deeper reasons* for actions, *i.e.*, things that

determine intentions to act. In what follows, I adopt this intuitive view. When the signer in Anscombe's example comes to know what their motive was, they don't learn whether their signing was intentional. Instead, they learn *why* they've signed intentionally in the first place.

Many things should be said about this understanding of motives—in particular, how exactly desires and further intentions relate to motives.[1] For brevity, I'll confine myself to one fundamental point: Are motives reasons or causes? Consider how 'motive' is used in a detective story. Crucially here, a motive isn't something that is necessarily acted on. The gardener can have a motive to murder, but it may have been the butler who did it. In *that* context, a motive is more like a 'pure rationalization', expressing that it *would* have made sense for the agent to act, independently of whether they actually did. Generalizing from this context, you might think of motives as unconnected to causality.

However, the sense of motive relevant for *knowledge* of one's motives is *not* this one. When the signer in Anscombe's example comes to know their motives, they don't just learn that it *would have made sense* for them to act out of admiration. Instead, they know that admiration *in fact* played some part in bringing about their action. Knowledge of your motives is incompatible with an agnostic stance about what actually made you do it. This shows that a motive, in the sense relevant here, entertains an *actual causal connection* to whatever it is a motive for.[2]

Having a somewhat clearer grasp of what motives are, we can now ask what it takes to know them.

2. TWO NECESSARY CONDITIONS

We don't need to think of reasons and causes as separated by an insurmountable schism (Davidson 2001). Dropping this assumption allows to appreciate that a motive-ascription is *falsified* if the cited motive has *no* causal bearing on the action. If admiration played no causal part in the signer's action, their belief that it influenced them would be false and thus couldn't amount to knowledge. Thus, the following seems like a good candidate for a necessary condition on knowledge of one's motives:

[1] I do this in more detail elsewhere (Hubacher Haerle ms).

[2] Non-causal theories can't explain *coming to know* one's motives (cf., Davidson 2001). Thus, I am committed to causalism about action-explanation (see §2). However, this is perfectly compatible with the view that any causal explanation *presupposes* a rationalizing explanation; we can think of 'pure rationalizations' as *constraining* the space of causal action-explanations.

(C1) A motive-belief which amounts to knowledge picks out *a* causal factor of an action.

As an absolute minimum, a motive-ascription expressing self-knowledge needs to refer to some causal factor in an action. Obviously, this is merely necessary and by no means sufficient: there are countless causes of any action that have nothing to do with motives at all.

Moreover, we know from epistemology that true belief isn't enough for knowledge. Instead, we need to have some form of *justification*. For the time being, I'll think of justification as imposing a sort of *anti-luck* condition (Pritchard 2007). Whether this is spelt out in terms of a reliable belief-forming process or along different lines (e.g. evidentialism), doesn't matter for my purposes. So, we can formulate a second necessary condition:

(C2) A motive-belief which amounts to knowledge is justified, *i.e.*, true in a non-lucky way.

Accordingly, we don't know our motives, if the belief we have only tracks a causal factor by accident.

Next, I will present research paradigms which have been taken to imply that it is not possible to fulfil those necessary conditions.

3. PSYCHOLOGICAL EVIDENCE

In a seminal article Richard Nisbett and Timothy Wilson (1977) discuss multiple studies where agents are manipulated into choosing certain items. The subjects were influenced in their choices by unconscious biases induced through the experimenters. When asked why they acted, agents came up with motives that had nothing to do with the true causes of their actions. From this, Nisbett and Wilson conclude that "people may have little ability to report accurately on their cognitive processes" (ibid.: 247). Instead, they argue that we may be able to infer them based on causal theories which can be at most "incidentally" correct (ibid.: 233, 253ff.).

Daniel Wegner and Thalia Wheatley (1999) advance a related argument. Drawing on a range of experiments—including the famous trials by Benjamin Libet—, they argue that people's actions are caused by unconscious mechanisms which *also* cause the feeling of conscious will, including the experience of acting out of a specific motive. However, this experience doesn't indicate an actual causal factor in an action; instead, it's a causally irrelevant 'third variable' (ibid.: 482f.; see *Figure 1*). The unconscious mechanism is what's really driving the action, the motive is merely 'epiphenomenal' (Thrash *et al.* 2010: 330).

An explicit successor of these research programs is the choice blindness-paradigm (Johansson *et al.* 2005; 2006; Hall *et al.* 2010; 2012). In the choice blindness framework, subjects choose between two alternatives. Afterwards, they are asked to justify their 'choice'. Unbeknownst to them, they are given the *other* alternative. Surprisingly, participants easily come up with reasons for the alternative they didn't choose. Now, those avowed motives couldn't have been causally relevant for their choice, because they justify the thing they didn't choose!

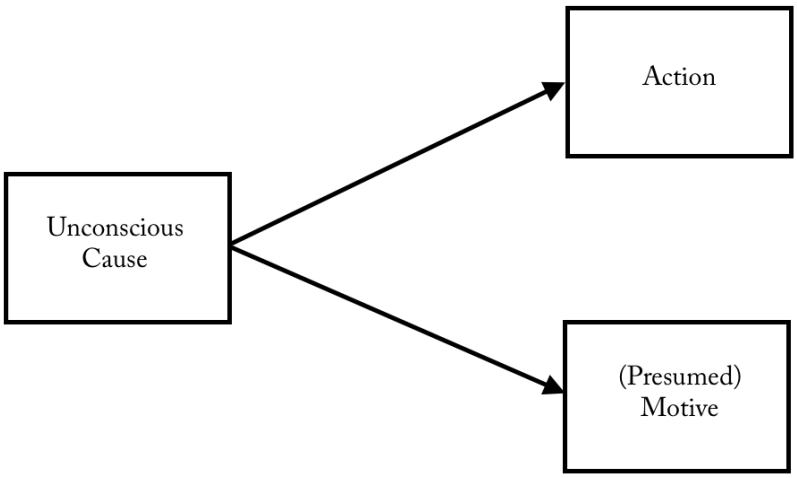

Figure 1: The unconscious cause leads to the action. The presumed motive has no causal bearing on the action.

4. SCEPTICISM

It's not hard to see how this evidence has been taken to motivate scepticism about knowledge of our motives. According to the *strong* sceptic, the evidence is best explained by the postulate that motives don't exist. We fail to meet (C1) because whatever we *can* consciously avow has nothing to do with the true causes of our actions. Some see Friedrich Nietzsche as a predecessor of this view (Ridley 2018).

More recently, psychologist Nick Chater has defended a related position when he claims that "the interpretation of the motives of real people is no different from the interpretation of fictional characters" (2018: 4). According to Chater, motive-beliefs can't be true as their content doesn't *exist* in the first place. Instead, our actions are determined by highly context-dependent processes (ibid.: 109ff.; cf., Steward *et al.* 2006). Uncovering a person's true motives is impossible "not because they are *difficult* to find, but because there is *nothing* to find" (2018: 4).

In the current literature, Chater's position represents an extreme and many writers favour a weaker position. Wilson (2002), John Doris (2015) and Hugo Mercier and Dan Sperber (2011; 2017) embrace something like *weak* scepticism about knowledge of one's motives. On their picture, the evidence isn't best explained by denying that motives exist. Rather, motive-beliefs are *systematically skewed*. Accordingly, the weak sceptic allows the occasional motive-belief to pick out an action's true cause, particularly when the action stems from conscious deliberation. But on that picture, there will be many situations where it *seems* to the subject that an action is caused by conscious deliberations, where in fact it stems from an unconscious cause. Often, what we take to be truthful ascriptions of motives are nothing more than confabulations.

This creates a sceptical problem (Scaife 2014: 480). It puts into question whether the processes producing beliefs about our motives satisfy the anti-luck condition. For simplicity, assume a *reliabilist* version of (C2): in order to satisfy (C2) we need a reliable process connecting our motive-beliefs to the actual causes of our actions (cf., Paul 2012). If such a reliabilist constraint holds, there will be a level of detachment between our motive-beliefs and the actual causes of our actions where even a true belief about our motives fails to amount to knowledge.

Now, I think there is ample reason to believe that many weak sceptics have a picture of the mind where the processes generating our motive ascriptions are not reliable. First, the weak sceptics' emphasis on the systematic disconnect between what we believe about our motives and the actual causes underlying our actions (e.g. conscious thought vs. adaptive unconscious), by itself suggests such a conclusion. If our motive-beliefs are so systematically detached, it's extremely plausible to think that the mechanism for ascribing motives doesn't reliably track causal influences on our actions, *i.e.*, has a reliability below 0.5 (Goldman and Beddor 2021). Following C2, however, it would need to do so in order for motive-beliefs to amount to knowledge.

Second, many weak sceptics hold that our practice of ascribing motives to ourselves didn't *evolve* to trace causes of our actions. Instead, this practice fulfils a narrative function where we create ourselves (Wilson 2002: 207), express our values (Doris 2015) or manage our reputation (Kurzban 2010; Mercier and Sperber 2011; 2017; Simler and Hanson 2018). This again suggests a sceptical attitude. One would feel pressed to come up with alternative functions of motive-ascriptions, only if one also believes that they do not reliably fulfil their apparent function.

Third, fulfilling any of these alternative functions will often entail failing to accurately report on the true causes of behaviour. For instance, always telling the truth about one's motivations is terrible advice for anyone seeking high regard among their peers. Further, values can be expressed, independent of articulating actual causes of actions. Of course, it could be that these alternative functions happen to incidentally also track true causes of behaviour. But it's unclear why we should believe so. Instead, it is much more plausible that *more often than not*, telling a narrative, managing my reputation or expressing my values will involve avowing motives that do not latch on to the true causes of my actions.[3] This, too, suggests that weak sceptics take the processes generating motive-ascriptions to be unreliable.

Even those authors who don't deny that we can occasionally get it right, hold a view which implies that (C2) isn't met: ultimately, we're too unreliable to *know* our motives.[4] Hence, the weak sceptic is a sceptic after all.

In the next section, I will argue against all forms of scepticism about knowledge of one's motives.

5. AGAINST SCEPTICISM

Scepticism about our ability to know our motives is a costly position. For a start, it's deeply implausible. According to the sceptic, when you talk to your friends about your true motives behind your choice of career, you are not speaking about what moved you to become a philosopher instead of an investment banker. Instead, you are doing something completely different, such as argumentatively justifying your actions (Mercier and Sperber 2011: 69), expressing your values (Doris 2015: 143ff.), constructing a fictional narrative (Wilson 2002: 15f.; Chater 2018: 4ff., ch. 6) or preserving your status (Kurzban 2010; Simler and Hanson 2018). Following the sceptic, we'd need to radically reconceive our practices of self-reflection.

Second, such a reconception would have far-reaching consequences. Many deeply important aspects of self-knowledge won't be available anymore. For instance, various forms of morality place great emphasis on the motives for which an action

[3] Thanks to an anonymous reviewer for pressing me to clarify this point.

[4] There is some reason to think that this argument carries over to internalist theories of justification too. Most internalists are committed to take evidence as a *means* to truth (Kelly 2016: §3) that's closely connected to *conscious experience* (Conee and Feldman 2008). It's questionable whether, on the picture offered by the weak sceptic, what we ordinarily draw on in order to know our motives can be considered evidence in that sense.

was conducted. Christian and Kantian moral thought holds that an action ought to be performed with good motives. Subjective consequentialism licenses your action as morally good, if your motive involved a rationally warranted belief that your action would maximize utility. If we can't know our motives, we have—at least on these pictures of morality—no way of learning about the moral character of our actions (cf., Doris 2015).

Another reason for the importance of knowledge of motives is that it is intimately linked to knowledge of one's values, goals, and character (Tiberius 2002: 159). Such knowledge is important, among other reasons, because it lets you reliably predict your own behaviour and facilitates collaboration with others (Baumeister 2011; Leuenberger 2021). If we can't know our motives, our self-knowledge is severely limited and we may be unable to predict, control and explain ourselves.

Clearly, the fact that a view has radical and far-reaching consequences is no argument against it. However, the stakes associated with a certain view can be relevant for its justification (Stanley 2005). A position as costly as scepticism about knowledge of one's motives needs extremely good reasons to be convincing. In the remainder of this section, I aim to show that there aren't any.

5.1. REPLICATION AND JUSTIFICATION

First, you might try to avoid scepticism by rejecting the evidence it's based on. Many results from social psychology have failed to replicate (Open Science Collaboration 2015). Likely this also affects the studies discussed above (§3). After all, they predate awareness of the 'replication crisis', rely on notoriously small samples and employ outdated statistical methods. Thus, there is an easy answer to the sceptic: simply deny that the evidence is any good.

However, to my knowledge, no failures of replication of the cited studies have come forth.[5] Thus, I will work with the assumption that the presented evidence is to be taken seriously (cf., Bird 2021). Should it turn out to be faulty, even worse for the sceptic. Until then, let's consider other replies.

In interpreting the evidence, it's important to note that in some sense, test subjects in Nisbett-and-Wilson-style studies are *right*. After all, they manage to give *justifying reasons* for their

[5] Instead, there are multiple papers defending the studies in Nisbett and Wilson (1977) against irreplicability (Guerin and Jones 1981; Sprangers *et al.* 1987) and the choice blindness experiments *do* replicate (Taya *et al.* 2014; Lachaud *et al.* 2022).

actions. Simply asking subjects "Why did you choose this?" leaves undetermined whether the question is after reasons or causes. Many interpretations of the empirical evidence tread on an equivocation between the two. What test subjects fail to do is to accurately report on their action's *causal history*. But they still manage to give *justifying reasons* (Keeling 2018; 2021: 324f.; Ganapini 2020; Andreotta 2021; Levy 2022). If being able to justify one's actions was sufficient for knowledge of motives, the evidence would fail to support scepticism.

However, this would leave us with something unrecognizable as knowledge of one's motives. As argued in §1, when the signer comes to know their motives, they learn more than that they can justify their action with appeal to admiration. They themselves concede that admiration isn't a good reason! Claiming that knowing justifying reasons is the same as knowing one's motives goes against (C1) and must be seen as a cop-out. We need yet another reply to the sceptic.

5.2. THE MYTH OF THE ONE TRUE MOTIVE

The inference from the psychological evidence to scepticism about knowledge of our motives rests on a mistaken assumption which I call 'the myth of the one true motive'. Here's how Donald Davidson (2001: 18) talks about false motive-beliefs:

> [Y]ou may err about your reasons, particularly when you have two reasons for an action, one of which pleases you and one which does not. […] You may be wrong about which motive made you do it.

Of course, Davidson is right in pointing out that we can be wrong about our motives. Yet, Davidson's quotation suggests that there is only *one* motive that 'made you do it'.[6] However, there seems to be no reason to think that an action can only have *one* cause. What speaks against multiple motives working together to cause an action?

First, you may think it incompatible with how we ordinarily think about causality, *i.e.*, as a relation with exactly two relata. It's true that we might intuitively adopt a picture of causality where each effect has exactly one cause. But, allowing for one and the same effect to have multiple causes, is much more adequate in a complex world. Modern scientific practice in biology or economics supports this (Kincaid 1996; Woodward 2000). The causal relationships under analysis there are far more com-

[6] While this passage *suggests* that, the idea that an action can have multiple causes is easily integrated into a Davidsonian picture.

plex than those assumed in a simple 'billiard-ball' picture. Indeed, analysing complex causal networks is the bread and butter of modern-day scientists. Therefore, a narrowly mechanistic framework of causality is a bad reason for believing in the myth of the one true motive.

Second, you may believe in the myth for reasons connecting back to the ethical theories sketched before (§5.1.). Recall the deontic idea that an action's moral properties are determined by its motives. The myth of the one true motive ensures that as long as your motives are morally definitive, your actions are as well. Clearly, we want to know if our actions are good or bad. For, it seems that our moral worth turns on it. Accepting the myth of the one true motive satisfies a desire to be morally definitive. However, on a realistic outlook it's clear that many actions are pervaded by morally laudable *and* blameworthy motives. And so are we as agents. Most of us have no definitive moral character. Instead, we are inherently *ambivalent* (cf., Velleman 2005; Kurzban 2010; Strohminger *et al.* 2017).

Thus, I think the myth of the one true motive is a by-product of a naïve notion of causality and a misguided wish for moral definitiveness. If we reject the myth and realize that motives over- and jointly determine actions, this changes how we can interpret the psychological experiments cited above (§3).

5.3. REVISITING THE EVIDENCE

Consider again a subject in a in Nisbett-and-Wilson-style-study. They are ask to choose between two alternative, they make their choice and say afterwards 'I chose this item because it looked the nicest to me'. If we accept the myth of the one true motive, the finding that subjects are reliably influenced by right-side bias rules out the possibility that the motive they articulate played any causal role in their action.

However, if we allow for co-determination between non-rational influences and motives, this inference isn't legitimate anymore. As Zina Ward and Edouard Machery (2018) argue, intentional motives and subpersonal effects can *both* causally contribute to our actions. Such causal over- and joint determination puts into question whether the participants in Nisbett-and-Wilson-style-studies are really *wrong* in their motive-ascriptions. It's true, they failed to track *all* the influences of relevant choice effects. Yet, that doesn't mean that the reasons given had *no* causal bearing on their action. A subject in a study may truthfully utter 'I chose this item because it looked the nicest to me' and this may be knowledge of their motives despite them *also*

being significantly influenced by a right-side bias (cf., Andreotta 2021: 4869). Being under the influence of choice effects doesn't mean you fall short of (C1), nor (C2); there's not *one* thing that made you do it (see *Figure 2*).

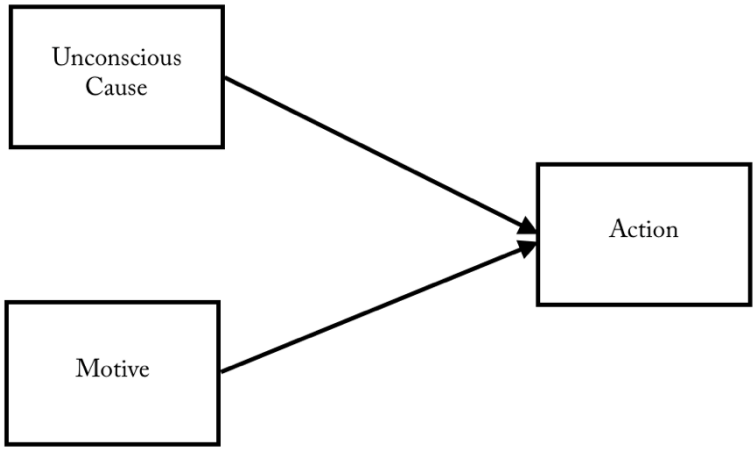

Figure 2: Unconscious cause and knowable motive overdetermine (or jointly determine) an action.

Rejecting the myth of the one true motive reveals that the experimental evidence underdetermines whether we should prefer a sceptical interpretation along the lines of *Figure 1*, where there is *no* causal connection between motive and action, over a non-sceptical interpretation represented in *Figure 2*, where there *is* a causal connection between motive and action.

Of course, it could be that if test subjects were given more time they would actually choose a different alternative or none at all. But even that doesn't challenge the claim that *in the moment of choice* the item they chose looked the nicest to them. Without the myth of the one true motive, the hypothesis that their report is truthful is at least as plausible than the hypothesis that they just made up that reason after the fact.[7]

Following the framework of Wegner and Wheatley, you might argue for an 'omitted variable' driving both the motive-ascription and the action and that's why we can't regard the motive as causally relevant (see *Figure 1*). Also here, the rejection of the myth of the one true motive reveals an alternative reading. It could just as well be that the unconscious effect causes the action *through* the motive (cf., Thrash *et al.* 2010: 323ff.). To return to the example used before: a right-side bias

[7] Thanks to an anonymous reviewer for asking me to clarify this point.

could make an object look especially appealing and this misleading appearance could trigger in the subjects a motivation to choose said item. On such an interpretation, it's not the case that the articulated motive and the actual cause for action stand in a reversed causal ordering. It's rather that the unconscious cause is just 'further up' the causal chain than the avowed motive (see *Figure 3*). In this case, however, there's no reason to assume that a subject's utterance of the form 'I chose this because it looked the nicest to me' isn't expressing *knowledge* of their motives. As David Lewis (1986) argues, for a causal claim to be true it doesn't need to capture an event's *whole causal history*. It would be absurd to expect someone to cite all causes of their action as their motive. Accordingly, I can know my motive, even if that motive is itself brought about by an unconscious cause.

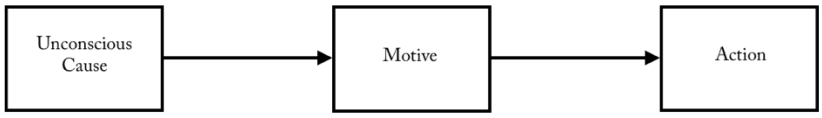

Figure 3: Unconscious cause causing the motive which in turn causes the action.

Let's apply this reasoning to an example given by Wilson (2002: 106f.):

> Suppose, for example, we observe a customer in a fast-food restaurant ask for a chicken sandwich, and we ask her why she ordered what she did. She would probably say something like, "Well, I usually order the burger, fries, and shake, but I felt more like a chicken sandwich and unsweetened iced tea today. They taste good and are a little healthier." These are precisely the thoughts she was thinking before she asked for the sandwich and thus were responsible for what she ordered—a clear case of conscious causality.
>
> Or is it? Suppose that earlier in the day the fast-food customer encountered someone who was quite obese, which primed issues of weight and self-image, which made her more likely to order food with less fat and calories than the burgers, fries, and shake. The customer was aware of part of the reason she ordered what she did—her conscious thoughts preceding her action—but unaware of what triggered these thoughts.

I'm unsure whether Wilson intends this to illustrate that the customer has no access to her real motives. Plausibly, her encounter with the obese person rendered certain facts about the

food options especially salient to her. On the basis of these salient features, the customer makes her choice. In that case, there is nothing wrong with her explanation of why she did what she did. The only possible charge would be that it's incomplete. But all explanations are. Wilson concedes that actions are unlikely to be "pure" and that she is aware of "part of the reason" why she acted (ibid.). My suggestion is that this is enough to know your motives. To ask for more would be—in a lot of cases—asking for too much.

Again, I'm suggesting that the psychological evidence underdetermines whether we should prefer a sceptical interpretation along the lines of *Figure 1*, where there *is* no causal connection between motive and action, over a non-sceptical interpretation represented in *Figure 3*, where there *is* a causal connection between motive and action. Accordingly, there is no compelling reason to assume that we are unreliable in forming beliefs about our motives. It's not clear whether participants in such studies really fail to know their motives; we have no grounds to deny that their beliefs meet (C1) or (C2).

To be clear, even if we adopt such a picture, we *can* read those studies in a way that does support scepticism about knowledge of one's motives. However, if we reject the myth of the one true motive, a non-sceptical interpretation of the psychological evidence becomes equally plausible. This undermines scepticism without treading on the ambiguity between causal explanations and rational justification. Now, scepticism about knowledge of one's motives starts to look unmotivated, especially given its far-reaching consequences.

What about choice blindness, though? My argument rests on the idea that the empirical studies give us no reason to disregard what subjects avow as part of the causal chain behind their action—either through over- and joint determination, or by referring to a link 'higher up' in the chain. Yet, clearly, this reasoning isn't available when it comes to choice-blindness. There, causal factors and avowed motives must be completely distinct.

I propose to question how much the choice-blindness findings generalize. Bear in mind how manipulative the experimental contexts are: the participants are consciously *tricked* and fail to track that trickery. Importantly, many ordinary life circumstances are not like that; we are highly unlikely to be manipulated in that way outside of an experimental setting. It's not clear to me that we can infer anything about our general abilities to know our motives from the fact that we are tricked in a highly-deceptive environment. Thus, the *external validity* of

the choice blindness results isn't obvious (cf., Sullivan-Bisset and Bortolotti 2021).

Still, the studies surveyed here only represent a small subset of the literature. Therefore, I make no claims about other experiments supporting scepticism about knowledge of one's motives.[8] What I've argued is that the evidence base for scepticism is smaller than often assumed. The burden of proof is thus on the sceptic to convincingly present evidence for why we should adopt their position. Until that happens, I conclude that scepticism about knowledge of one's motives ought to be rejected. What the evidence shows is how easily we are manipulated into forming motives on mistaken beliefs (e.g., that some of the alternatives *are* better than the others). What it doesn't show is that we have troubles *knowing* these motives—on any sensible conception of what such knowledge means.[9]

6. CONCLUDING REMARKS

In this paper, I've issued a warning against taking certain psychological studies as proof that we can't know our motives. I've shown how costly such a position is and that if we reject the myth of the one true motive, it starts to look unmotivated.

Nonetheless, I don't want to deny that knowing one's motives is difficult. Scepticism is partly so attractive because knowing your motives is hard. For much of our motives run unconsciously. Uncovering them is no easy thing to achieve. Moreover, we have a strong desire to present ourselves in a flattering manner, even to ourselves. This desire makes it likely that when we think about our motives, we are engaged in motivated reasoning skewing our self-inquiry (Kurzban 2010; Simler and Hanson 2018; Williams 2020). If we want to know our motives, thus, we need to keep these two factors in mind and their distorting effects at bay.

If knowledge of one's motive is hard, don't we fail to meet (C2)? This worry is exacerbated if we note how weak (C1) and (C2) are. On the picture offered here, it's possible to know your motives even if all you're reliably picking out is an *insignificant* cause. You could think that knowing something which matters very little to how you act can't constitute knowledge of your

[8] I do take my argument also to apply to the studies led by David McClelland (Thrash *et al.* 2010) and Michael Gazzaniga (Wilson 2002: 99ff.; Chater 2018: Ch. 6).

[9] It may well be that these findings raise problems for other aspects of self-knowledge, such as first-person authority (Scaife 2014; Levy 2022). However, they don't' attack the *possibility* of knowing our motives.

motives. Therefore, you might be in favour of a further necessary condition:

(C3) A motive-belief which amounts to knowledge picks out a *relevant* causal factor of an action.

If (C3) is in place, scepticism seems more plausible again. If anything, the evidence shows that the most relevant causes are often hidden from our view; position effects have a much higher predictive power than the motives people can articulate.

I'm not completely convinced by (C3). Perhaps all we can do in trying to know our motives is registering factors that are, causally speaking, pretty minor. Still, there might be a lot of value in that. But even if we do add (C3), there are more ways of avoiding scepticism.

First, it's not clear that difficulty and unreliability are as closely connected as the objection assumes. Certain activities—such as climbing a hard mountain route—can be difficult without often provoking failure. Instead, their difficulty is grounded in the amount of *effort* needed. Knowledge of one's motives might be hard because of the care it requires, and not because of frequent failure.

Second, even if psychological evidence established the unreliability of *one* process of introspection, this isn't the only one. In the cited studies the participants were mostly forced to rely on solitary introspection. They had little time and no dialogue partners available. As a result, the methods of self-inquiry they could draw on were severely confined. In real life, however, there are a lot of methods we can use in trying to know our motives. You can talk to your friends, go to therapy, write into your diary, meditate, etc. (Vazire and Mehl 2008; Thrash *et al.* 2010). Any of *these* processes are likely to be more reliable than sole introspection under a heavy time-constraint.

So, even if we make the conditions for knowledge of one's motives more demanding, there are good reasons to think that they can be met. In some sense, knowing your motives is a lot like *science*. Both in science and in introspection, we need to draw on a variety of methods to gather a trustworthy evidence base. Also, we need to take great care in establishing and solidifying theories about different causes and how much they matter. I develop the idea of coming to know one's motives as a 'science of the self' elsewhere (Hubacher Haerle ms).

Knowledge of your motives is like many valuable things in life—it requires a great deal of effort. With respect to a Kantian obligation to know one's motives this means that the attack

fuelled by recent psychological evidence can be deflected. Fulfilling our duty to know our motives may be quite hard, but it is *not* impossible.[10]

(5'873 words)

BIBLIOGRAPHY

Andreotta, A. (2021), "Confabulation does not undermine introspection for propositional attitudes", *Synthese*, 198: pp. 4851–4872.

Anscombe, G. E. M. (2000 [1957]), *Intention*, Harvard University Press.

Baumeister, R.(2011), "Self and identity: a brief overview of what they are, what they do, and how they work", *Annals of the New York Academy of Science*, 1234: pp. 48–55.

Bird, A. (2021), "Understanding the Replication Crisis as a Base Rate Fallacy", *The British Journal for the Philosophy of Science* 72:4: pp. 965–993.

BonJour, Laurence. *The Structure of Empirical Knowledge*. Cambridge, Mass: Harvard University Press, 1985.

Chater, N. (2018), *The Mind is Flat*, Yale University Press.

Conee, E. and Richard F. (2008), "Evidence", in Smith, Q. (Ed.), *Epistemology: New Essays*, Oxford University Press.

Davidson, D. (2001), *Essays on Actions and Events*, Oxford University Press.

Doris, J. (2015), *Talking to our selves: Reflection, ignorance, and agency*, Oxford University Press.

Ganapini, M. (2020), "Confabulating Reasons", *Topoi*, 39: pp. 189–201.

Goldman, A., and Beddor, B. (2021), "Reliabilist Epistemology", in Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*.

Guerin, B. & Innes, J. (1981), "Awareness of Cognitive Processes: Replications and Revisions", *The Journal of General Psychology*, 104: pp. 173–189.

Hall, L., Johansson, P., Tärning, B., Sikström, S. & Deutgen, T. (2010), "Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea", *Cognition*, 117(1): pp. 54–61.

[10] I'm indebted to Richard Holton, Jessie Munton, Tatiana Sitnikova, Dan Williams and an anonymous reviewer for invaluable feedback on this paper. Ancestors of it have benefitted from comments by Rhea Blem, Demetra Brady, Angela Breitenbach, Joane Marner, Michael Müller, Jordan Scott and Paulina Sliwa. Also, I am thankful to Michał Barcz, Paula Keller, Will Hornett, Cecily Whitely, Kyle van Oosterum and Aiden Woodcock and audiences in Cambridge and Vienna for helpful conversations of this material. This research was supported by the Boustany Foundation and the Swiss Study Foundation.

Hall L., Johansson P., Strandberg T. (2012), "Lifting the veil of morality: choice blindness and attitude reversals on a self-transforming survey", *PLoS ONE*, 7(9), Pe45457.

Hubacher Haerle, P., "Knowing Your Motives", unpublished manuscript.

Johansson, P., Hall, L., Sikström, S. & Olsson, A. (2005), "Failure to detect mis- matches between intention and outcome in a simple decision task", *Science*, 310: pp. 116–19.

Johansson, P., Hall, L., Sikström, S., Tärning, B. & Lind, A. (2006), "How something can be said about telling more than we can know: On choice blindness and introspection", *Consciousness and Cognition* 15(4): pp. 673–92.

Kant, I. (2006 [1785]), *Groundwork of the Metaphysics of Morals* trans. by Gregor, M., in Wood, A. (Ed.), *Immanuel Kant: Practical Philosophy*, Cambridge University Press.

—(2006 [1797]), *The Metaphysics of Moral*, trans. by Gregor, M., in Wood, A. (Ed.), *Immanuel Kant: Practical Philosophy*, Cambridge University Press.

Keeling, S. (2018), "Confabulation and rational obligations for self-knowledge", *Philosophical Psychology*, 31: pp. 1215–1238.

— (2021), "Knowing our Reasons: Distinctive Self-Knowledge of Why We Hold Our Attitudes and Perform Actions", *Philosophical and Phenomenological Research*, 102: pp. 318–341.

Kelly, T. (2016), "Evidence", in Zalta, E. (Ed.), *The Stanford Encyclopedia of Philosophy*.

Kincaid, H. (1996), "Causes, Confirmation and Explanation", Ch. 3 of *Philosophical Foundations of the Social Sciences*, Cambridge University Press: pp. 58–110.

Kurzban, R. (2010), *Why Everyone (Else) Is a Hypocrite : Evolution and the Modular Mind*, Princeton University Press.

Lachaud, L., Jacquet, B., Baratgin, Reducing, J. (2022), "Choice-Blindness? An Experimental Study Comparing Experienced Meditators to Non-Meditators" *European Journal of Investigation into Health, Psychology and Education*, 12: pp. 1607–1620.

Leuenberger, M. (2021), "What is the Point of Being Your True Self? A Genealogy of Essentialist Authenticity", *Philosophy*: pp. 1–23.

Levy, Y. (2022), "Neo-Ryleanism About Self-Understanding", *Inquiry*.

Lewis, D. (1986), "Causal Explanation" in *Philosophical Papers 2*, Oxford University Press: pp. 214–40.

Mercier, H. & Sperber, D. (2011), "Why do humans reason? Arguments for an argumentative theory", *Behavioral and Brain Sciences*, 34: pp. 57–111.

—(2017), *The Enigma of Reason*, Harvard University Press.
Moore, C. (2015), *Socrates and Self-Knowledge*, Cambridge University Press.
Nietzsche, F. (1954 [1888]) *Twilight of the Idols*, trans. Kaufmann, W., Viking.
Nisbett, R. & Wilson, T. (1977), "Telling more than we can know: Verbal reports on mental processes", *Psychological Review*, 84: pp. 231–59.
Open Science Collaboration (2015), "Estimating the reproducibility of psychological science", *Science*, 349.6251, aac4716.
Paul, S. (2012), "How we know what we intend", *Philosophical Studies*, 161: pp. 327–346.
Pritchard, D. (2007), 'Anti-Luck Epistemology', *Synthese* 158: pp. 277–97.
Ridley, A. (2018), *The Deed is Everything: Nietzsche on Will and Action*, Oxford University Press.
Scaife, R. (2014), "A problem for self-knowledge: The implications of taking confabulation seriously", *Acta Analytica* 29(4): pp. 469–485.
Schultheiss, O. and Brunstein, J. (2010), *Implicit Motives*, Oxford University Press.
Simler, K. & Hanson, R. (2018), *The Elephant in the Brain: Hidden Motives in Everyday Life*, Oxford University Press.
Sprangers, M., van den Brink, W., van Heerden, J., Hoogstraten, J. (1987), "A constructive replication of White's alleged refutation of Nisbett and Wilson and of Bem", *Journal of Experimental Social Psychology*, 23(4): pp. 302–310.
Stanley, J. (2005), *Knowledge and Practical Interests*, Oxford University Press.
Stewart, N., Chater, N. & Brown, G. (2006), "Decision by sampling", *Cognitive Psychology*, 53(1): pp. 1–26.
Strohminger, N., Knobe, J., & Newman, G. (2017), "The true self: A psychological concept distinct from the self", *Perspectives on Psychological Science*, *12*(4): pp. 551–560.
Sullivan-Bissett, E., & Bortolotti, L. (2021), "Is choice blindness a case of self-ignorance?", *Synthese*, 198(6): pp. 5437–5454.
Taya F, Gupta S, Farber I, Mullette-Gillman OA (2014), "Manipulation Detection and Preference Alterations in a Choice Blindness Paradigm", *PLOS One*, 9(9): e108515.
Tiberius, V. (2002), "Virtue and Practical Deliberation", *Philosophical Studies*, 111: pp. 147–172.
Thrash, T.M., Cassidy, S.E., Maruskin, L.A. & Elliot, A. J., (2010), "Factors that influence the relation between implicit

and explicit motives", in Schultheiss and Brunstein (2010): pp. 308–346.

Vazire, S., & Mehl, M. R. (2008), "Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior", *Journal of Personality and Social Psychology*, *95*(5): pp. 1202–1216.

Velleman, D. (1989), *Practical Reflection*, Princeton University Press.

— (2005), "Identity and Identification" in *Self to Self*, Cambridge University Press: pp. 330–360.

Ward, Z. and Machery, E. (2018), "'Defeaters' don't matter", *Behavioural and Brain Sciences*, 41: pp. 52–53.

Ware, O. (2009), "The Duty of Self-Knowledge", *Philosophy and Phenomenological Research* 79 (3): pp. 671–698.

Wegner, D. M., & Wheatley, T. (1999), "Apparent mental causation: Sources of the experience of will", *American Psychologist*, 54(7): pp. 480–492.

Williams, D. (2020), "Socially adaptive belief", *Mind & Language*, 36: pp. 333–354.

Wilson, T, (2002), *Strangers to ourselves*, Belknap.

Woodward, J. (2000), "Explanation and Invariance in the Special Sciences", *British Journal of the Philosophy of Science*, 51: pp. 197–254.