The Moral Status of Automatic Behaviour

by
Sebastian Rodriguez Duque

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Carleton University
Ottawa, Ontario

© 2017, Sebastian Rodriguez Duque
Abstract

*Automaticity is pervasive in human behavior. Our theories of normative ethics assume a model of human moral psychology that does not take this into account. If much of human behavior proceeds automatically, then we need to ask what we should make of the moral status of such behavior. I argue that we can take automatic behavior to be morally assessable, that is, blameworthy or praiseworthy, in its own right. First, because there are features of automatic behavior in virtue of which it proceeds intelligently. Second, because acts of deliberation are not necessary for acting for reasons.*
Acknowledgements

I would like to thank my supervisor, Myrto Mylopoulos, whose boundless patience and insightful advice made the writing of this thesis a delightful and rewarding process. I would also like to thank Annie Larivee, whose mentorship and enthusiastic support have made my pursuit of graduate work possible.
# Table of Contents

- **Abstract** .................................................................................................................................. 1  
- **Acknowledgements** .................................................................................................................. 2  
- **Table of Contents** ..................................................................................................................... 3  
- **1 Chapter: Introduction** ............................................................................................................. 5  
  - 1.1 Introduction ............................................................................................................................. 5  
  - 1.2 Setting Up the Problem ........................................................................................................... 13  
- **2 Chapter: Automaticity and Skilled Action** ............................................................................. 23  
  - 2.1 Characterizing Automaticity. .................................................................................................. 23  
  - 2.2 A Working Definition. .......................................................................................................... 26  
  - 2.3 Skill ...................................................................................................................................... 28  
- **3 Chapter: On Consciousness in Action** .................................................................................... 36  
  - 3.1 The Functional Role of Consciousness in Action Production and Control ......................... 36  
  - 3.2 Consciousness and Intentions ............................................................................................... 37  
  - 3.3 Action Guidance .................................................................................................................... 39  
  - 3.4 Action Planning ..................................................................................................................... 48  
- **4 Chapter: Automaticity in Moral Psychology** ....................................................................... 54  
  - 4.1 Gendler’s Aliefs ..................................................................................................................... 54  
  - 4.2 Moral-Worth and Belief ........................................................................................................... 57  
- **5 Chapter: Automaticity and Acting for Reasons** .................................................................... 66  
  - 5.1 Deliberation ............................................................................................................................ 66  
  - 5.2 Acting For Reasons ................................................................................................................ 70  
  - 5.3 Desire .................................................................................................................................... 72
1 Chapter: Introduction

1.1 Introduction

Research in psychology and neuroscience is strongly suggestive of the prevalence of automatic behaviour in human agency. Combined with philosophical considerations that question the role of deliberation in acting for reasons, this offers a new and complex picture that is far removed from much of our folk theoretical assumptions about the nature of human moral psychology. Moral psychology is the branch of philosophy which concerns itself with describing the psychological underpinnings of moral agency to ensure that normative ethical theorizing is not at odds with the most up-to-date scientific evidence and reflects the psychology of real agents as closely as possible. As such, empirical research that illuminates the mental processes at work in generating a given human behaviour is of direct relevance to normative ethical theorizing.

There are several ways in which we can ask the question of just what it means for an agent to be a good agent. Most broadly, we can ask the axiological and metaethical questions surrounding what, if anything, goodness itself may turn out to be. The range of realist and anti-realist positions suggest a number of possibilities as to how to tackle our

construal of goodness as a property or value we assign to ourselves or an action we perform. They raise the ontological and epistemic questions of whether goodness even exists, and if it does, whether we can ever get to know it.

Another way to ask this question about just what it means to be a good agent is provided by normative ethics as we commonly understand it. Some of its main representatives within the Western tradition include Kantian deontology, consequentialism, virtue ethics, or feminist care ethics. These theory traditions posit the agent as reason-determining, and using various standards of value by which to measure the appropriateness of their actions. Usually, the reason-determining process involves some kind of conscious deliberation, either publicly or privately, in at least some stage of the reason-determining process.

By “reason-determining” I only mean the process by which an agent would discern the right reason for which to act when faced with a morally relevant situation. Of course, not all situations are morally relevant. Perhaps, one can argue, along some model of virtue such as Zen Buddhism, that all actions do in fact include a prudential dimension that is of a broader ethical relevance. And while it may be true that a view about our morally relevant psychological make-up defended here could be compatible with such a view, it is not a view for which I need to argue, and which I will therefore not consider further. Either way, even if all actions turn out to be ethically relevant in some broad axiological construal, it is still not the case that all actions are morally relevant.

A final way to ask the question about just what it means to be a good agent is provided by moral psychology. Moral psychology is the branch of philosophy that, drawing from empirically informed research in fields such as experimental psychology,
neuroscience, and cognitive science, broadly construed, poses the question of just what it is that an agent is up to when we take them to have undertaken a morally relevant act.

I am working from the assumption that a theory about morality (or anything) should at least not be fundamentally incompatible with the best available science. We should at least not posit the existence of a flying spaghetti monster that issues moral dictums as it hovers menacingly over our atmosphere as the source of all morality. On this assumption, we should take the absence of scientific evidence about such a creature hovering over the Earth’s skies as good enough reason for us to never seriously countenance it as a possibility. Moreover, when it comes to compatibility with theories like the theory of natural selection; theories which are patently well supported, we should take it that our philosophical theorizing ought to be compatible with these. Thankfully, much of what science has given us remains inconclusive, which leaves both scientists and philosophers with plenty of room for debate. But the point is that when it comes to a project such as supplying the explanation of just what constitutes human consciousness and action, and how this might bear on our construal of morality, then a mutually informed approach seems like the most sensible.

Moral psychology can bear directly on just what metaethical theory we might find most attractive, or with what we think a theory of normative ethics should hold. If it turns out that humans are cognitively impaired to such an extent that we can only cognize about events within the physically bounded realm, then at least epistemically speaking, we need to rule out any possibility of our accessing any possible supra-natural property of goodness, no matter the ontological matter of fact. Likewise, in normative ethics, if it turns out that automaticity is prevalent, then we may need to revise our theories that rely on an explicit
consequentialist calculus, or the consideration of how to apply the categorical imperative to a given situation.

To proceed, however, this project takes for granted the division of labour described above going forward. I take it for granted that there is something such as moral action, and that there are features of action in virtue of which one might decide whether an action is or is not suitably morally assessable. I therefore also assume that there is something such as goodness, or badness, and that we can decide whether an agent is worthy of either of those labels. I will argue that what grounds the praiseworthiness or blameworthiness of our moral actions is acting for reasons, and what I intend to produce is an alternate account of acting for reasons that goes beyond the traditional assumptions of folk psychology (FP). This, in turn, paves the way for automatic behavior to be morally assessable.

I argue that the theoretical assumptions underpinning much of normative ethical theorizing must be heavily revised considering the available empirical evidence. More specifically, normative ethics seem committed to a problematic set of assumptions about how agents undertake right action, and, most importantly, pin the moral assessibility of action on those grounds.

Let me unpack this by first illustrating what I am referring to. Consider, for example, the approach advocated by Kant where the agent must deliberate about prospective action choices such that she can determine the best end, an end which must be compatible with the categorical imperative. Or, in consequentialist theories at large, where the agent must deliberate about what to do by way of a cost benefit analysis vis a vis a given final value, be it goodness, happiness, or pleasure. Or, in Aristotelian virtue ethics, where the incipient virtuous agent must deliberate so that she can approach as best as
possible the mean between the vices she is trying to avoid. In these theories, the usual assumption is that after deliberating about what to do, the moral agent undertakes the action as a matter of their volition or their will. The moral agent deliberates about what to do, and then forms an intention which forms the basis of her action. This is the reason for which she acts. And, this is the reason that we can judge to be right or wrong. It is the locus of the moral assessibility of her action.\(^2\) Of course, we can also judge whether the agent deliberated in good faith, or if her acting for the right reason was sincere. However, if she deliberated in bad faith, then presumably her reason for action will be wrong. If she acted based on an ostensibly good reason insincerely, then perhaps we can think that her real reason for action was a deception, and that her purported reason for action was only apparently good, or beneficial to others in some fashion, but still not worthy of moral praise in relation to the agent herself. What we are after then, is reasons for action that are praiseworthy in relation to the agent herself. This will be important later in my discussion of automaticity. I will argue that, like acting on purportedly good reasons insincerely, acting in a seemingly praiseworthy manner automatically will only be praiseworthy if it is done in virtue of the right reasons as they relate to the agent herself.

In moral psychology, the divide of this construal for what it means to act for reasons is most obvious within the traditional rationalist versus sentimentalist debate. Broadly, between processes that involve conscious, deliberative decision making, or processes that are grounded on other cognitive features such as our emotional appraisals of a given

\(^2\) Although in consequentialism the rightness or wrongness of an action does not depend on deliberating correctly, I take it that no calculation of utility/consequences could come without a deliberative process. The action would be judged as moral based on its adherence to that expected consequence, not the deliberation, but deliberation is nevertheless an integral part of the process.
situation. However, I argue that this sort of divide is grounded more deeply in our folk psychological assumptions about the nature of human agency, as illustrated by the assumed central role of deliberation in acting for reasons. This set of assumptions, as Ellen Fridland makes the case, creates a dichotomy between the conscious/controlled and the non-conscious/automatic, and maps it onto intelligent versus unintelligent, to use Fridland’s terms. I will follow Fridland, Bargh, Moors, and De Houwer in problematizing this division in order to present a picture of moral agency that is much more complex. As Heidi Maibom points out, while much of the recent empirical evidence seems to favour the sentimentalist side of the picture, the evidence in fact suggests the need to reconceptualise these boundaries.

This first introductory chapter will lay out the debate to which I hope to contribute, in broad terms. By reflecting on certain cases, I will motivate the need to wonder about what moral assessability and acting for reasons might look like if they are not to be grounded on intentions or conscious acts of deliberation. Setting up the negative part of the project, I will present what Michael Brownstein and Alex Madva (henceforth B&M for the sake of brevity) have termed the Top Down Harmony (TDH) view, which in general states that a behaviour is morally praiseworthy or blameworthy iff it accords with our consciously

---

4 Fridland, “Automatically Minded”; “Skill and Motor Control”
6 Maibom, “Experimental Evidence in Moral Judgment”
endorsed beliefs. This will be the target view that I will seek to undermine, to make the case in favour of intelligent, moral automatic action.

Starting with Chapter 2, I will provide a working definition of automaticity and examine automaticity from the point of view of skilled behaviour. I will argue, first, that automaticity is pervasive, and second, that recent accounts of skilled action provide grounds to think that automaticity proceeds intelligently. As a related, but not integral point to this section, I will argue in the final chapter that, as a possible way forward, skilled action is a good place to look for accounts of moral behaviour in light of the evidence.

After characterizing automaticity, and providing grounds for why we might think it proceeds intelligently, I will then address what I take to be three central sources of resistance from our folk theoretical assumptions for thinking that automaticity might be morally assessable.

First, from FP, we might think that automaticity is not morally assessable because it is not guided by conscious intentions, and such guidance is necessary for moral assessability. In Chapter 3 I examine the available evidence for thinking that consciousness controls behaviour, and call into question the strength of that folk psychological assumption. Chapter 4 will present Tamar Gendler’s and then B&M’s accounts of automaticity within moral psychology. I will discuss their two differing versions of how to assess the moral worth of actions when they are mediated by automatic processes. I will

---

disagree with the view that B&M attribute to Gendler, the Top Down Harmony view (TDH), and expand on the cases which I touch on below, such that we can think that automatic actions are morally praise or blameworthy in their own right.

The fifth chapter will move on to examine Nomy Arpaly and Timothy Schroder’s (henceforth A&S) work on non-deliberative action. I will present A&S’s work to undermine our FP notions that acts of deliberation are necessary to act for reasons. The final, sixth chapter, will provide a discussion of the previous views, with the aim of suggesting interesting ways forward in how to construe moral agentive behaviour. My argument will be that moral action viewed as moral expertise, an idea already comfortable in the virtue ethics tradition, will be strongly reflective of what empirical evidence can tell us about how human moral psychology proceeds. In the end, we should have strong grounds to think that automatic behaviour is morally assessable in its own right. I will have addressed three central sources of resistance from what I take to be our folk assumptions: 1. That conscious intentions are necessary for guiding behaviour; 2. That moral assessibility is grounded on the accordance of our actions with our avowed beliefs; and 3. That morally assessable behaviour needs to be behaviour done for reasons, and acts of deliberation are required for acting for reasons. It will emerge that 1. We cannot be very confident about the role of consciousness in action production or guidance at this time; 2.

---

Coherence with our avowed beliefs is not necessarily tied to moral assessability; and, 3.
Acting for reasons happens in virtue of processes other than deliberation.

1.2 Setting Up the Problem

I begin this section by setting up three sample cases that can give us pause in order to
motivate the discussion that follows. The only aim of the following examples is to highlight
the range of automatic behaviours that can arise, and why we might wonder how, while at
least two of them seem morally assessable, it is not clear to what we would attribute the
source of this assessment.

1. Nomy Arpaly and Timothy Schroeder recount an anecdote about Oscar
Wilde, “who according to one story bragged that he could make a pun on any
subject. Presented with the challenge “the queen,” he quickly answered, “the queen
is no subject.””

2. B&M draw our attention to the awkward and potentially disastrous occasion
of Obama’s swearing in ceremony in 2009 when Chief Justice John Roberts
stumbled over the wording while administering the oath of office. As they tell it,
“[w]ith a slightly puzzled brow, Obama hesitated before repeating Robert’s botched
phrasing, then smiled widely and nodded slightly to Roberts, as if to say, “It’s okay,
go on.” As they interpret it, the spontaneous nonverbal cues from the president

---

10 Ibid., 229.
11 Brownstein and Madva, “Ethical Automaticity.” 75.
12 Ibid.
put his automatic social adeptness on display as he smoothly navigated and relieved a potentially difficult situation.

3. B&M give us the example of Wesley Autrey, “who saved the life of an epileptic man who, in the midst of having a seizure, had fallen onto the tracks of the New York City subway. Autrey jumped onto the tracks himself and held the man’s body down while the train passed just inches overhead.”

The central feature of the three examples in the list above is that they are intended to showcase actions that are in some way admirable or praiseworthy. In the case of the second example, it is not to say that we have to like puns, but it is to say that we usually appreciate quickness of wit. Here is an action that on the face of it does not seem to be subject to a conscious decision or deliberation to act in such or such a way. And, it is praiseworthy because of this fact. As A&S point out, in no way would we think of Wilde as clever if he had gone home and pondered a suitable pun for a few days before returning to respond. The central theme this points to is the argument that will be brought forward by Fridland, and that we will examine in the following chapter. Namely, that automatic behaviour can be intelligent. It is in the juxtaposition of conscious/automatic, rational/irrational, intelligent/unintelligent, that we are losing track of cases that involve an entirely different set of phenomena that deserve examination. In this anecdote about Wilde

13 Ibid., 85.
we see automatic behaviour that is clever, and we see that it is clever in large part because it is automatic behaviour.

The final two examples from the list are the two which are meant to provide cases that are of moral relevance. In the case of President Obama, the implicit question of what he should do is not answered through his own personal deliberation on the matter. First, because it seems like there would be little time to do so, and second because I take it as uncontroversial that, although we can to a limited extent, we do not often give conscious micro-directions to our facial expressions. In fact, as we shall see later, it is precisely these sorts of finer grained responses which are carried out through automatic processes. Instead, the question of what Obama should do in that awkward and potentially disastrous scenario is answered, B&M want to say, by him spontaneously, through his engaged attunement to his situation.\(^{15}\) I will argue that this is in fact a case of automatic action, which is also a case of morally assessable, in this case praiseworthy, action. It will be my task to unpack what, if anything, is happening in this case which allows us to think of this action as moral.

The case of Autrey jumping in front of the tracks of an oncoming train to save a complete stranger is the most compelling. Normally we take a central feature of altruism to be selflessness, and we take altruism to be a hallmark of morally laudable behaviour. In cases of self-disregarding readiness to lay down one’s life for another human, we normally take this to be an exemplary action, and would normally consider it heroic. This example is notable because it is a real-life example of a person with no apparent training for what

\(^{15}\) Brownstein and Madva, “Ethical Automaticity,” 75.
to do in a moment of crisis responding in what we could consider to be the most laudable form of moral behaviour.

This is like the sort of willingness to jump into danger displayed by first respondents to an emergency, but with the important difference that these respondents are trained to do so. We can wonder what sorts of dispositions are at work for Autrey such that in a moment of crisis he acted the way that he did. A further, interesting detail about his case is that when questioned about it he was unable to explain his action. He had been travelling with his two children that day, and as B&M point out, upon reflection, if he had been asked whether he would do what he did while abandoning his two children, he would likely have wanted to at least think about it.16

The above three examples point to a range of instances of automatic behaviour. It is not clear from the last two examples whether they are in fact cases of moral behaviour yet, and what if anything, it is about the way those actions are carried out that would still make them out to be moral. For the actions to be called moral, it seems crucial that for Obama, in the first instance, his apparent smiling to alleviate the tense situation, was in fact because of a desire to alleviate the tension of the situation. Likewise, for Autrey’s action to be heroic, it seems crucial that he undertook it to save the stranger, and for no other reason. The final two examples, therefore, draw out another central theme of this thesis, namely, that the automatic actions in question, to be considered moral, need to be performed for the right reasons.

Moving on from these examples, a further aspect that needs to be emphasized in this discussion is the case of actions that are discordant with consciously avowed beliefs.

16 Ibid., 87.
Cases where agents hold ‘rationally’ determined beliefs; beliefs arrived at on the basis of some process of deliberation; and, actions which seem to completely disregard those consciously avowed beliefs. This is the case with, for example, implicit bias. B&M point to the well-documented phenomenon of ‘aversive racism,’ which is the practice of people who may sincerely believe themselves to be non-racists, and nevertheless exhibit behaviours which,

[make them] more likely to hire a white job candidate over an equally qualified black candidate, more likely to find a black defendant guilty than a white defendant with equally incriminating evidence, and more likely to exhibit a range of discriminatory “microbehaviours” (e.g., they tend to make less eye contact, make more speech errors, and sit further away from black interlocutors). 17

This discordance, and the effect of it on our actions is potentially alarming. For somebody like Gendler, this means that we must somehow ensure that our beliefs and our actions are in harmony with each other as often as possible, and it is easy to see why. I will partially unpack some of her view in the interest of offering it as a contrast to the set of views I will propose as the better explanation. I will return to Gendler’s view in order to do full justice to her helpful insights, in Chapter 4. Suffice it for now to say, however, that in order to explain the discordance between belief and action such as in cases of implicit bias, she posits a new mental category to explain the sort of causal antecedent of action which might be at work. She calls these ‘belief-discordant aliefs’, and she uses the concept of aliefs as a useful construct that can explain the sort of occurrences that we have been discussing. First, however, it is not clear how the introduction of a new concept like alief is useful. 18

Indeed, the proliferation of new concepts such as Schwitzgebel’s own contribution of

17 Ibid., 70
18 In conversation with Myrto Mylopoulos
‘beliefishness’\(^{19}\) does not seem entirely useful. And it is not useful mainly because, within psychology, as Gendler herself acknowledges, there is a massive amount of evidence on something already called ‘automaticity’, the multi-faceted nature of which does not seem to lend itself to a monolithic description, other than as a range of phenomena that share a set of common features at different times.

However, for the moment, the relevant issue is the explicit association between arational/associative/automatic that is at the heart of what she takes a belief-discordant alief to be. In fact, within the implicit hierarchy of what is reasonable, alief falls below rational belief, as an affect driven state more akin to mental states we might share with non-human animals. This is not to say that Gendler thinks that aliefs are always belief-discordant. Instead, she thinks that aliefs are, especially given research on automaticity, a pervasive aspect of most if not all action.\(^{20}\) However, the crux of the case, as already mentioned, boils down to this perceived tension between rational belief, and a-rational alief. Naturally then, for her, in cases of belief-discordant aliefs, the solution is to bring those discordant aliefs in line with our rationally produced beliefs.\(^{21}\)

So here is why, even if we keep the notion of aliefs alive, the target must be the implicit assumption that these mental events are somehow unintelligent. The case of implicit bias illustrated nicely the manner in which automatic actions can be entirely discordant with our consciously held beliefs. The second and third examples, those of Obama and Autrey, both illustrated cases of praiseworthy action where there was no easily

\(^{19}\) Schwitzgebel, Eric. "Acting Contrary to our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief," \textit{Pacific Philosophical Quarterly} 91 (2010): 531-553.

\(^{20}\) Gendler, "Alief and Belief," 642.

\(^{21}\) Ibid., 661.
distinguishable occurrent rational conscious process at work in the formation of a belief that might have been tied to an intention in order to produce an action. They were actions that happened in the absence of a conscious process of deliberation, and in the absence of an intention derived from that process. This leads us to my final example for this section, by Nomy Arpaly.\textsuperscript{22} The example is this: Huck Finn finds himself traveling down the Mississippi with Jim, an escaped slave. Despite his considered judgments that he ought to turn Jim in because he is an escaped slave, he never does, even though he has multiple opportunities to do so. Arpaly highlights Huck’s tortured deliberations on the subject; his determination to do the right thing as he saw it; his subsequent failure to live up to his own decision; and the guilt and remorse he suffered from his own view of himself as a morally bankrupt person. Arpaly’s main contention is that it is because of other background psychological mechanisms at work in Huck that he is able to do the right thing in spite of his own consciously endorsed judgment on the matter. These are the processes that A&S call non-deliberative, and which, because they can happen outside the control of consciousness or outside the direction of a specific intention that is a product of a process of deliberation, I take to be compatible with my general account of automaticity. This is the contrast case to that of implicit bias. It is an example of a belief discordant action which produced a morally laudable result. It is important to think about how we respond to the case of Huckleberry Finn. First, because it will help us get clear on our intuitions about the role of belief in moral action. One of these might be that our consciously held beliefs should matter quite a lot in deciding our correct moral course of action. A second one might be

that Huck was being quite a good person after all. And not just an accidentally good person, but a person whose real moral fibre shone through, despite his confused and mistaken beliefs. In that case, we must find a way to reconcile those two intuitions. Second, it is important to think about our response to the case of Huck because it provides grounds for us to think that certain kinds of automatic behaviour can be moral in their own right, independently of conscious attitudes or beliefs. This, in combination with the final two examples of our original list, provides us with a suite of cases which serve to motivate an inquiry into the normative status of automatic behaviour.

This leads me to my final point of this introductory chapter. While I will not be appealing to it, B&M accept the possible usefulness of a concept like alief to help explain cases of automatic behaviour.\textsuperscript{23} What they do not accept, however, is Gendler’s ultimate conclusion that our best solution is to ensure that our beliefs and actions need to be concordant in order for our actions to be appropriately moral. This is what they dub the Top-Down Harmony view (TDH) mentioned above. As they express it, “TDH is the view that an automatic disposition such as an alief is in good ethical standing if and only if it is governed by an agent’s considered beliefs and ends.”\textsuperscript{24} B&M seek to enrich the conception of the ethical standing of aliefs themselves by “[conceiving] the normative relations implicit in these interactions via analogy with the normative role commonly attributed to reactive attitudes.”\textsuperscript{25} As such they are “[a]utomatic affective responses [which] reflect long processes of enculturation and experience through which tacit feelings of right and wrong

\textsuperscript{23} Brownstein and Madva, “The Normativity of Automaticity,” 411.
\textsuperscript{24} Brownstein and Madva, “Ethical Automaticity,” 72-73.
\textsuperscript{25} Ibid., 77.
move agents to action.” These operate through the feelings of an agent reacting to the tension of a situation in a manner that will adjust behaviour in order to alleviate that tension. These “[f]elt tensions are marked by either positive or negative valence, which acts like a physiological reinforcer of possible behaviours.” Crucially then, and contrary to the TDH approach they attribute to Gendler, their “account makes no reference to the standing of aliefs relative to an agent’s considered beliefs or reflectively endorsed ends.” On their account, the occurrence of alief is more akin to a skill for which there can be no reflective belief to accord with. In short, it is a kind of know-how that does not need to make reference to the propositional content, or know-that, of beliefs in order for it to count as the relevant kind of knowledge that needs to be available in order to guide right action in certain situations. From this we can see how a view might begin to take shape, which allows automatic behaviour to be morally assessable in its own right.

In this chapter I have introduced the general themes that will be operative in this thesis. Most importantly I have addressed the perceived importance of the discordance between some automatic action and some conscious states. The view that an action’s ethical status is determined by its relationship to the propositional content of conscious states (beliefs), has been set up as the target view that I will seek to undermine throughout this thesis. With examples that made a case for assessable moral action that was nevertheless either belief-discordant, or simply unable to be subject to a belief, I have sought to motivate the discussion to follow. For now it is only necessary for the reader to

26 Ibid., 78.
27 Ibid., 79
28 Ibid., 80
29 Ibid., 81
think there are reasonable grounds in order to pursue an account of phenomena that would seem to have independent normative status from that of our avowed beliefs.
2 Chapter: Automaticity and Skilled Action

In this chapter I examine some of the current literature on skilled behaviour which can lead us to a more nuanced characterization of automaticity. In this way, we can begin to paint a view where, despite the prevalence of automaticity in agency, human agents can still be acting for reasons. At least, as we shall see in this chapter, there are reasons to think that skilled agents are acting intelligently all the way down to their fine-grained motor control. The chapters that follow will discuss the possible nature of moral agency bearing in mind what I first discuss here.

2.1 Characterizing Automaticity

That automaticity is pervasive is widely accepted in experimental psychology. As Bargh and Chartrand note, the consensus in the experimental psychology community is not an ‘either-or’ scenario, but rather, that a combination of automatic and controlled processes are involved in the production and control of agentive behaviour.\(^3^0\) One popular source of such characterizations has been Kahneman’s work, which posited a dual system theory based on an intuitive, automatic mode, along with an online, conscious control mode.\(^3^1\) This in turn has been eagerly taken up by psychologists interested in moral agency, some of whom have gone so far as to posit characteristically deontic and consequentialist types

---

\(^3^0\) Bargh and Chartrand, “The Unbearable Automaticity of Being.”

of judgments to the two streams, where the more emotional, intuitive judgments have been correlated with the automatic stream and associated with deontic judgments, and the purportedly effortful, reflective judgments have been associated with consequentialism, especially in the work of Joshua Greene and Jonathan Haidt.\textsuperscript{32} As I pointed out in the previous chapter, in philosophy, moral psychologists have pointed to the pervasiveness of automaticity to elucidate issues around moral responsibility, or implicit bias. Others have pointed to experimental results that purport to show that human character traits are fundamentally unstable, and are instead subject to the priming or triggering of automatic processes.\textsuperscript{33} Overall, as Tamar Gendler notes, automaticity has been broadly accepted and established in the psychological literature.\textsuperscript{34} This leaves the question of just what to make of this fact with regards to the moral status of such behaviour.

One good starting point is to wonder about how best to characterize automaticity. However, there is little conceptual clarity on just what automaticity amounts to. A first rough and ready characterization is to contrast automaticity with conscious control. As such, with regards to judgment, and specifically moral judgment, some authors in psychology use intuition or emotion, often interchangeably (as processes that may arise both subconsciously and automatically), and contrast them with rational thought and conscious control (as processes which arise both consciously and bear phenomenal


\textsuperscript{34} Gendler, “Ailief and Belief,” “Ailief in Action.”
characteristics of proceeding via conscious effort). However, this is not all that automaticity is meant to capture. In accounts of motor control, for instance, automatic processes are usually used to account for the control of finer grained motor control, which may be happening under the guidance of a broader goal representation. And, as I also noted, in moral psychology the idea of automaticity has been used to refer to actions that are discordant with our avowed beliefs, especially if the beliefs are occurrent.

Two of the most enthusiastic proponents of the role of automatic processes in moral judgment have been Greene and Haidt. Both authors have argued together and separately in favour of the prevalence of mechanisms other than slow, conscious, deliberation as responsible for the greater part of action. They present evidence from brain scans in order to support a view where moral judgment is tied to the automatic processes of the brain. However, while in his other work Greene often speaks about emotions as the mental states responsible for our behaviour, Haidt often speaks of intuitions. This can cause confusion, as it is unclear whether both authors are referring to the same processes with different terms, or if there is some more basic causal role that the term intuition is meant to capture, and where emotion is playing the descriptive role of the process at large. Either way, because both terms come with heavy theoretical baggage, I suggest we drop talk of either intuitions or emotions and for our purposes simply allow that in the end behaviour guided by either of these states falls in the realm of what we would call automatic. While their work is a related and important source for thinking about the nature of moral judgment, their questions are not the central focus of my project. I am not exploring whether moral

---

35 For a detailed discussion, see Maibom, “Experimental Evidence in Moral Judgment.”
36 Fridland, “Automatically Minded.”
judgments are generated automatically. Rather, I am exploring whether there are features of automaticity such that we can think it is morally assessable. Because of this, I will not consider Greene and Haidt’s work further.

2.2 A Working Definition

There are different literatures that use somewhat different ideas of what is meant by an automatic process. They encompass everything from higher cognitive states like emotional appraisal, down to the micro-movements of my pupils as they survey an interlocutor’s face. In fact, automaticity is a broad category with no single feature that is either necessary or sufficient to characterize the phenomenon.37

Fridland favours “the decomposable, gradual view of automaticity”38 defended by Moors and De Houwer. As she explains,

this conception of automaticity allows for a notion of automaticity to be circumscribed by a set of characteristics, but not necessary or sufficient features. These features will have their own time course for development and so will characterize various automatic processes to various degrees during different times in learning. This approach allows us to retain a concept of automaticity, which is both operationalizable and theoretically useful for guiding our thinking.39

Conceptualizing automaticity in this way allows us to entertain ideas about the workings of those processes that go beyond mere brute-causal accounts of behaviour that is not controlled. This makes sense; psychological processes are complex and multi-variate, with different processes being involved at all levels of behaviour. We can say that automaticity is pervasive, because several of these processes proceed automatically, but

37 Moors and De Houwer, “Automaticity, Analysis.”
39 Ibid., 10-11.
we cannot say that there is one or a definite set of features characteristic of all these processes. As Fridland continues:

> automatic processes are generally fast, parallel, efficient, effortless, resistant to stressors, relatively unaffected by cognitive load, do not require conscious guidance, are not the result of specific intentions to deploy them, etc. However, this does not mean that every automatic process will be equally characterized by each of these features at all times.  

I endorse the view of automaticity supported by Fridland. If we think back about Oscar Wilde in the previous chapter, then we can see how his wit qualifies as automatic under this definition. It is fast, efficient, effortless, and was not the result of a specific intention to deploy ‘wit’ – he simple acted. Likewise, the cases of Obama and Autrey bear many of these features. In the case of Obama, we can especially notice the absence of specific intentions for a given behaviour, when so much of the important content was resting on the efficiency of the micromovements of his face. Finally, in the case of Huck, while he was deploying specific intentions about what to do, and his deliberation had yielded a certain decision (that of turning in Jim), we can see that his actions are mediated by something other than this, and are instead undertaken in virtue of background processes that were at odds with his occurrent cognitions, and at odds with his conscious intentions. We might disagree that the case of Huck is indeed a case of automaticity, or that it is not quite like the other three examples I have presented. However, what is important about the Huck case is that he is acting for reasons outside of a process of deliberation, and this is

---

40 Ibid., 11
the point that is related to the moral assessability of automaticity which I will discuss further in Chapter 5.

2.3 Skill

Looking at accounts of skill can help us elucidate the different factors at play in automaticity at large, and how they might play a role in moral psychology. What is immediately interesting about skill is that while there is an increase in the level of automation of the skilled process as the agent becomes more experienced, there is also a corresponding increase in control. This is a helpful insight that is immediately available if one considers skilled performance in sport, or the skilled playing of a musical instrument. In these instances, the motor control of at least the finer grained movements is automated, and as this is increasingly the case, it is also the case that their performance is progressively better. Evidently, automatic processes are therefore not processes that necessarily proceed non-consciously (one is usually conscious of playing an instrument), nor are they processes that are not controlled.

Fridland argues that the distinction between automatic and controlled behaviour deserves re-examining considering recent evidence. She questions the usual dichotomy of the intelligent and unintelligent with which the controlled and automatic get tagged from the get go. Most often, those processes which are explicit and controlled are assumed to be ‘intelligent’ processes, whereas all those others which are automatic are assumed to be ‘unintelligent’.41 On those views, unintelligent processes are “unconscious, implicit,

automatic, unintentional, involuntary, procedural, and non-cognitive,” whereas intelligent processes are “conscious, explicit, controlled, intentional, voluntary, declarative, and cognitive”.

Usually, intelligent processes are taken as “processes that need to be cashed out in semantic or psychological terms – prototypically, propositional states that are conceptual, compositional, recombinatorial, generalizable, and that can enter into logical reasoning.” Unintelligent processes, on the other hand, “can be explained according to mechanistic, causal principles, without need to appeal to concepts, semantics, or personal-level intentional states.” At its root, then, the distinction is merely the familiar philosophical dichotomy “between the psychological and the physical/biological, the semantic and the causal, the mind and the brain.”

As Fridland points out, one illustrative example is Dreyfus. Dreyfus’ account of skill is through the role of emotion. For him, the embodied, somatic markers of emotion are at the root of the experience which accrues over time into skilled expertise. As Fridland points out, “on Dreyfus’s account, expert skill is a case of non-deliberative, non-reflective, arational, atheoretical action… In Dreyfus’s hands, the automaticity of expert action highlights its opposition to the thoughtful, calculating, deliberate, controlled, conscious

42 Ibid.
43 Ibid.
44 Ibid.
45 Ibid.
46 Ibid.
qualities of rational processes." This account by Dreyfus is illustrative because it involves the bias of a skilled process that involves automation being juxtaposed to rational processes. In addition, the automation is cashed out in terms of emotions, which is a rather narrow subset of what might be thought of as automatic. It is not clear why we should privilege emotions as the primary drivers of automation.

The central element of Fridland’s account of intelligent automaticity is that we can think automatic processes are cognitively penetrable. On her view, automatic processes are cognitively penetrable because they are semantically linked to task relevant, personal level intentional states. Fridland notes that

[T]he action-control system is automatic in two ways: it is triggered automatically with the formation of the appropriate intentions and it runs automatically under the guidance of those intentions… That is, the agent need not deliberately, consciously, intentionally, or effortfully initiate the working of the action-control system. The system is initiated automatically as a result of the formation of strategic intentions; it is not initiated as a result of the intention to initiate the system.49

In her 2016 paper, Fridland reviews the accounts of motor skill as presented by Papineau,50 and Stanley and Krakauer.51 They are both, on her view, overly simplistic accounts of what motor skill amounts to. They cash out the motor component of skill as reflex-like, or in terms of motor acuity, both of which are construed as bottom up, brute-causal processes. Fridland isolates 5 features of motor skill that should hold true if these accounts are correct. If they are correct then “motor control is: (1) ballistic, (2) invariant,

49 Ibid., 17.
(3) independent of general action trajectories, (4) insensitive to semantic content, and (5) independent of personal-level intentions.\textsuperscript{52}

However, evidence in favour of optimal control theory suggests that on all these counts, skilled motor control is other than what those authors should expect, based on their account. On the count of ballisticity, where automatic processes are cashed out in terms of brute-causal, bottom up events that are essentially reflex like in nature, Fridland points to the evidence of skill itself. The evidence is that as skill increases, so does an agent’s level of control. Papineau characterizes skill as being composed of both a cognitive element and a brute causal, reflex like element. Undertaking the right skill is a matter of the right cognitive state activating the right dispositional folder for the attendant motor dispositions to then unfold. As Fridland points out, if they are construed as mechanistic processes independent of the cognitive content of the intentional states that set them off, then we should also expect to find them resistant to change or modification. However, evidence she points to, by for example Logan,\textsuperscript{53} on the motor processes of skilled typists, shows the ability of typists to stop or intervene in the unfolding of their skilled behaviour. In other words, once set off, the purported dispositional folder did not then proceed inevitably to unfold as programmed, but rather, was susceptible, and responsive, to changes in the intentional content of the individual.

Likewise, when it comes to an expected increase in invariance of learned motor processes, the evidence points to the contrary. One piece of evidence by for example

\textsuperscript{52} Fridland, “Skill and Motor Control,” 2.
Todorov and Jordan,\textsuperscript{54} is that, as it turns out, the particulars of movement are variable down to their finer grained iterations. As a matter of fact, flexibility in adjusting to different circumstances is a marker of expertise. If dispositional action folders or practiced motor acuity were simply just deployed and then carried out rigidly as learned, expert individuals would be clumsy and unable to adapt to changing circumstances. Moreover, further evidence indicates that variability is decreased to a greater degree in those parts of movement that are related to task-relevant success. From this we can conclude both that variance does not decrease into what becomes a rigid, reflex-like process, and that the variance is directly responsive to task relevant success, such that it can alter or stabilize depending on the circumstances. As this evidence suggests, and in relation to point three (3) of the list above, the general trajectories of automatic processes are also not independent from other higher cognitive states. They are in fact flexible and changeable in virtue of them, that is, in direct relation to their task-relevance.

The crucial question is then, how it is that automatic processes remain sensitive to task-relevance or irrelevance, if the unfolding of automatic processes is not happening in virtue of pre-selected causal pathways. For Fridland,

\begin{quote}
the motor control system is able to differentiate between task-relevant and task-irrelevant dimensions of movement by having the capacity to respond directly and flexibly to task goals. Accordingly, and as optimal control theory holds, the ability to differentiate between task-relevant and task-irrelevant details of a movement is accomplished by a sensorimotor control system that is directly sensitive to the semantic content of personal-level goals.\textsuperscript{55}
\end{quote}

So that in effect, the motor component of the control system does not proceed in a brute-causal way once the appropriate intention has been activated, but rather, adjusts to

\begin{footnotesize}
\begin{tabular}{l}
\end{tabular}
\end{footnotesize}
achieve task success in each setting. In this way, the final implication that Fridland attributes to other versions of hybrid theories of motor control is addressed. Namely, given the sort of responsiveness evident through the cognitive penetrability of personal-level intentions onto motor processes, it is clear that automatic process do not proceed unintelligently.

Further to the motor control system, there is another area where fine grained control has been shown to be susceptible to cognitive penetrability: selective attention. This kind of attention is responsible for selecting the relevant features on an action space to which an agent must attend in order to gather the relevant information, given the agent’s goals, plans, and strategies. Importantly, this kind of attention improves with training, is often automatic, and is directly sensitive to the semantic content of intentional states at the personal-level. Agents need not consciously or effortfully direct their attention. Rather, once a personal level intention for action has been activated, then the selective attention becomes engaged. As we will see from the experiments of Kentridge et al. in chapter 4, selective attention can proceed both consciously and non-consciously. In this way, and as Fridland notes, “agents cannot always deploy selective attention deliberately because skilled agents often do not know which features they attend to in skilled performance.” This way, selective attention is especially interesting because even when it is not deployed deliberately, it is still often sensitive to the semantic content of personal level intentional states. In this way, “for automatic, selective attention to function effectively, it must be sensitive, not simply to the existence, but to the content of the reasons, goals, and strategies that the agent

---

57 Ibid., 19.
possesses at the personal-level." In support of this, both Fridland and Wu refer to the experiments by Yarbus where the movement of participants’ eyes as they surveyed a painting “responded in a systematic way to the distinct questions they were being asked.” In the painting, four people are seated facing a door, where a fifth person has just walked in. The participants were asked how long the visitor had been gone, and to try and determine the age of those in the painting. As Fridland explains, “when it comes to telling someone’s age, the face is a great place to attend. When trying to figure out how long someone has been away, the expressions of greeting between people is a better indicator.” For Fridland this is evidence that automatic processes are cognitively penetrable.

For my purposes then, there is a clear account for us to think that automatic behaviour is proceeding intelligently, and more importantly, it provides grounds for developing an account for acting for reasons through automatic behaviour. That is to say, on this account, acting for reasons is cashed out in terms of the personal-level intentions that Fridland has in mind. However, that Fridland cashes out the cognitive penetrability in terms of a top-down view much like the TDH view attributed to Gendler falls short of doing justice to the evidence. While it is clear that automatic processes do not proceed in a brute causal way irrespective of cognitive states, the evidence we have already looked at in terms of belief discordant actions provide grounds for us to wonder whether this commitment to personal-level intentions by Fridland is on the right track. Specifically, if we think of cases which might not be subject to an intention, as some of the cases in the

38 Ibid.
39 Wu, Wayne. “Experts and Deviants.”
62 Ibid.
next chapter arguably are not, and if, further, we think about Huck Finn, and how his automatic actions (as actions done in the service of ND processes that are not directed by specific intentions) were directly at odds with his personal level intentions, then we might wonder if there is more to the picture. I will argue that on A&S’s account of acting for reasons, we have a way of cashing out cognitive content that is relevant for penetrating our automatic behaviours, but which is not necessarily acquired in virtue of personal-level intentions of the sort Fridland seems to have in mind. The final chapters develop this account. However, what is most important for this chapter is simply the idea that there is evidence which shows that features of automaticity are intelligent and are responsive to the substantive sort of content which would make up a moral reason. In this way not only can we be comfortable with the pervasiveness of automaticity, but we can continue the discussion by beginning to countenance that there are features of automaticity in virtue of which an action might be morally assessable. While we need not commit to an account of moral action as a form of skilled behaviour, we can think that skilled behaviour is a good place to look for the sort of evidence that shows automaticity to be a nuanced and subtle category.
Chapter: On Consciousness in Action

3.1 The Functional Role of Consciousness in Action Production and Control

In this chapter I address what I take to be a central feature of our folk theoretical commitments about action. This is meant to tackle the first of the three central FP assumptions I have attributed to normative ethics, namely, the assumed role of conscious guidance in action. As we shall see the functional role of consciousness in agency remains unclear. While there are reasons to doubt empirical evidence that was once taken as a robust challenge to the existence of conscious mental causation, problems remain for theorists who would save a functional role for consciousness in exercises of agency. I aim to point to what seems like a crucial requirement if we wish to preserve a functional role for consciousness in agency. For such a role to remain, it seems likely that something about action production and control should happen in virtue of a conscious state’s being conscious. This seems obvious but, as I will point out below, arriving at that conclusion also seems far from easy. In view of this I discuss the account defended by Pacherie, where she aims to preserve a role for consciousness in agency by unpacking different features of intentions. I will make the point that while her account of intentions is compelling, the place she reserves for the functional efficacy of conscious states falls short. This is because no facet of action production or control seems to occur in virtue of a

---

63 For scepticism that there is such a role, see David M. Rosenthal, “Consciousness and its function,” *Neuropsychologia*, 46, no. 3 (2008): 829-840.
conscious state’s being conscious. And, while Pacherie points to instances where conscious states are present in the unfolding or planning of agency, it is not clear how their being conscious affects the process of action production or control itself. I argue that a full explanation of conscious agency must not only point to instances where conscious states are present during the process of action production, but also point to what happens in action production and control in virtue of those conscious states’ being conscious. For my purposes, this means that the features of automaticity stressed in Chapter 2 are prevalent, and need to be taken into consideration by theories of normative ethics. In turn, this also points to the inadequacy of the TDH view espoused by Gendler, as to the necessary role of anything like conscious avowal in order for an action to be assessable as moral.

3.2 Consciousness and Intentions

In modern experimental psychology, the causal efficacy of consciousness in action production has been questioned at least as far back as the now famous Libet experiments.\(^6\) Libet seemed to demonstrate that action readiness potential—the neural activity thought to be the signature of action initiation—was apparent in the brain well before any conscious intention to act was manifest, as reported by subjects. In other words, the directly antecedent mental states for an action being undertaken (as understood by the brain getting ready to perform said action), were non-conscious mental states. The conscious ‘decision’ came after the brain was already showing signs of action initiation. This was largely taken

to show that what counted as our conscious decisions were in fact preceded by various non-conscious processes which bore direct causal responsibility for action initiation.

As Pacherie explains, these results have led to an ongoing debate with various consequences, not the least of which is what Bayne and Levy have termed ‘will skepticism’, where various philosophers and psychologists have used the evidence to fuel a general worry about the very possibility of freedom of the will. However, as Pacherie points out, this seems to sit comfortably with compatibilism. Moreover, as Pacherie also points out, we may have good reason to question the evidence itself. For example, recent studies have taken on the core assumption that the ‘readiness potential’ is what is in fact being observed in the EEG scans. Instead of showing the brain ‘ramping up’ before the execution of an action, these recent studies have explained the EEG ‘preparatory activity’ as part of just so much random noise as is constantly appearing in scans of brain activity.

In his studies, Libet found that readings of readiness potential preceded “subjective estimates of the time of the conscious ‘urge’ to move by 300ms or more.” However, according to two new studies, “the apparent build-up of this activity, up until about 200 ms pre-movement, may reflect the ebb and flow of background neuronal noise, rather than the outcome of a specific neural event corresponding to a decision to initiate movement.” Instead, it is due to the artificial time-locking feature of the Libet studies themselves that there is the appearance of a connection between these spikes in activity and the moment a

---

69 Ibid.
decision is made.\textsuperscript{70} In light of these considerations, it is reasonable to turn again to the possible functional role of consciousness in agency, and try and explain just what it might be. Moreover, as the next section argues, it may be that the mistake is to focus solely on action initiation in the first place.

3.3 Action Guidance

Pacherie divides her analysis into what happens both before and after action initiation. This is a further reason why, along with questioning the prevailing empirical account, we might still be able to accord a functional role to consciousness in agency. In this way, even if the Libet experiments prove to be right and the immediate causal antecedents of action initiation are indeed non-conscious processes, there is still plenty of room for conscious states to play a role. I will address what happens before action initiation in the following section, for now, let me focus on what happens after.

Following Harry Frankfurt, Pacherie points out that causal theories of action go wrong when they focus on the moment of action initiation alone, and forget to account for what happens while the action itself is being performed. Frankfurt argues that in action control, agents are still appropriately causally related to the unfolding of an action insofar as they can adjust its course, or control whether the action unfolds at all.\textsuperscript{71} This is referred to as action guidance. On Pacherie’s account of intention, the nature of action guidance seems to provide a place for consciousness, as the guidance of an action involves both

\textsuperscript{70} Ibid., 78
conscious and unconscious states, and controlled and automatic processes. It is the fact that these conscious states are present, and the features of action guidance they seem to facilitate at the time of their presence, that should give us reason to ascribe to them a functional role in action guidance. However, as I will argue, upon closer scrutiny, it is not clear what the functional role of those conscious states *qua* conscious states is, even if they are present. All aspects pointed to by Pacherie where a conscious state might be present can also be undertaken by unconscious mental processes, and it is not clear what, if anything, is happening in virtue of a state’s feature of being conscious. On her account, an array of mental states that can be said to be part of an intention, perform the job of action guidance irrespective of whether they are conscious or not. This leaves open the question of the relevance of consciousness to action guidance.

It is right to shift our focus away from the mere moment of action initiation and instead focus on what happens after and before to explain the role of consciousness in the production of action. One attractive feature of this move is that, as Frankfurt explains, it allows us to avoid issues of causal deviance that arise within the causal theory of action. Roughly, the causal theory of action counts a behaviour as an action, as opposed to a mere happening, so long as it has as its direct antecedents appropriately related causal states. These states are causally efficacious in the production of the ensuing action, and on some views, they are what we often call an intention.

In at least our folk theoretical assumptions, it is here that the locus of conscious activity in the production of an action would seem to reside. That is, through a rudimentary account of action, it can be assumed that it is as a matter of conscious brain activity that a

---

72 Ibid., 157
decision to undertake an action, an intention, has been formed, and it is this intention which is appropriately causally related to the ensuing action. Of course, it is here that issues around causal deviance, or evidence such as that of the Libet experiments can be presented to throw this sort of account into doubt. In cases of causal deviance, the intended outcome for an action is achieved, but it arrives through a divergence from the intended causal pathway. To paraphrase Frankfurt’s own example, a criminal is supposed to spill their drink to signal the trigger for whatever crime should be committed by them and their team, but they are so nervous their hand’s shaking spills their drink. As a result, the crime is committed as intended, but we would be hard pressed to say that the spilling of the drink was causally related in the appropriate way to an intention to undertake the crime.73 Frankfurt is therefore motivated to resolve these issues of causal deviance by pointing out that causal theories of action have neglected the very target of their theory: the action itself. He offers a theory of action guidance which then accounts or agency in terms of the control an agent is able to exert over an action as it unfolds, irrespective of its causal past.

Building on Frankfurt’s overall point about control and guidance, Pacherie proposes a view that uses the internal model theory of motor control and her own dynamic view of intentions in order to explain agential control, and thus provide a place for conscious agency within action guidance. Briefly, the internal model theory of motor control posits both forward and inverse models of control. As an action unfolds, the inverse models compute representations of finer grained motor control needed to undertake a given movement. Efference copies of these representations are sent to forward models, which predict the sensory consequences of their execution, and the representations outputted by

the inverse models and the forward models are then compared to ascertain whether adjustments in the trajectory of the movement are needed.\textsuperscript{74} Of course, as Pacherie points out, this says nothing of conscious agency yet. The internal model, and action control as envisioned by Frankfurt, can both very well proceed non-consciously. However, coupled with the view of dynamic intentions that Pacherie defends elsewhere,\textsuperscript{75} the different representations at work in the unfolding of an action provide room for a role for conscious states.

In her dynamic view of intentions Pacherie posits a hierarchy where representations are divided into three levels.\textsuperscript{76} Importantly, these are porous, non-serially occurring divisions which have much more to do with levels of specificity than with causal order. However, what this allows for is levels of broader goal representations to be active consciously, even while non-conscious, automatic processes are happening in tandem. However, we can re-phrase this claim as saying that it allows for automatic, non-conscious processes to take their course even while conscious broader goal representations are active in tandem. The relevant question is whether the presence of the conscious broader goal representation indicates that some feature of action guidance is happening in virtue of them.

Pacherie points to a set of experiments where, “participants were instructed to move a stylus on a graphic tablet along a straight line to a visual target. Participants could not see their drawing hand, but its trajectory was visible as a line on a computer screen. On some trials, the experimenter introduced a directional bias electronically so that the visible

\textsuperscript{74} Pacherie, “Conscious Agency,” 38.
\textsuperscript{76} Pacherie, “Conscious Agency,” 38.
trajectory no longer corresponded to that of the hand.”77 While the bias remained small, the subject performed adjustments to the trajectory of their hand non-consciously in order to continue to draw the straight line towards the visual target. However, “it is only with larger biases that participants became aware of a discrepancy and began to use conscious correction strategies to compensate for the bias and reach the target.”78 According to one of the experiments by Fourneret et al. cited by Pacherie, schizophrenic patients performed more poorly as compared to controls when it came to making the necessary corrections to their movements.79 The relevant difference, according to the experimenters, is that, while the controls were consciously aware of a disconnect between their motor and visual inputs, the schizophrenic patients were unable to have conscious access to their motor inputs and consequently privileged their visual inputs in their attempt to correct their movements. The result was a noticeable difference in the effectiveness between the two groups. According to the experimenters, it would seem like the relevant difference between the two groups is the lack of conscious access to the mismatch between the motor-visual inputs. In other words, it is in virtue of this conscious access that the correction is able to take place, to the degree that it does. That is to say, it is a more precise correction; the schizophrenic subjects still effected a correction, it was just less accurate than the controls.

However, this is in contrast with an earlier set of similar experiments also by Fourneret et al., also cited by Pacherie, where what is highlighted is the difficulty of normal

---

78 Ibid.
patients in tracking their own movements. \(^{80}\) When asked to verbally rate the displacement of their hand, the normal subjects, “tended to grossly underestimate the deviation of their hand trajectory” \(^{81}\) with respect to its deviation from the straight line. Crucially, for both sets of experiments, subjects were presented with cards at the end of the trials in order to rate the level of deviation of their hand from the straight line. In the later set, the schizophrenic group performed poorly as compared to the controls, but in the earlier set, the normal patients were found to perform poorly overall. In light of these earlier findings, it is not clear what to make of the findings of both sets of experiments taken together. However, if what matters for the success of the control subjects in the first set of experiments I mentioned is their conscious awareness of a discrepancy between their motor and visual inputs, given these earlier (second) set of experiments it is not clear why the point to emphasize is the efficacy of the conscious awareness of either set of patients.

The crux of the issue with the above account is the following: there is a difference between awareness of agency, and agency itself. It is beyond the scope of this paper to review different accounts of the phenomenology of agency. \(^{82}\) Suffice it to say that consciousness, and a general awareness of agency are at the root of what these questions are about. The seeming discrepancy between this awareness of agency, and what can be explained as our actual agency, is what is at issue. Therefore, in the above experiments, the question is not whether the subjects of the experiments were able to self report accurately


\(^{81}\) Ibid., 1138

about a correction after the fact. It is whether the correction of the trajectory of the line was corrected in virtue of an awareness of the directional bias introduced into the experiment. It is not clear why we should extrapolate from the above experiments that the relevant awareness is that of a disconnect between the motor-visual inputs, especially in light of the earlier findings. While the evidence that schizophrenic subjects were less aware of their own movement, and that control patients were better at performing certain aspects of the corrections, is suggestive of a role for awareness of motor inputs, this does not constitute direct evidence that corrections were happening in virtue of such an awareness.

This is more so the case given the fact that the report of conscious awareness kicks in after the fact in the experimental design, when subjects are asked to rate their own hand movement. What seems clear from these experiments is not that patients were taking conscious correction strategies. Rather, they show that as the corrections got larger given the greater degree of perturbation, they became more accessible to conscious monitoring. Crucially, conscious monitoring here does not need to imply control. What the experimenters set out to do was investigate the ability of our conscious awareness to track bodily movement. First, the findings seem to show that we are not very good at it. Subjects by and large seemed not to be consciously aware of the corrections they were performing to the trajectory of their pen, unless it was from a large perturbation. And second, even in those few instances where subjects reported being aware of their bodily movement, a further claim is needed in order to show that the corrections were happening in virtue of this bodily awareness. Taken together, all these experiments seem to be reporting is the limited ability for conscious awareness of agency to take hold even while undertaking
corrective bodily movements, and that control patients have a lower threshold for conscious awareness about a motor-visual mismatch of information.

Moreover, the experiments are resting on an assumption about the nature of the conscious awareness involved in following visual cues. If the correction of the trajectory can be explained via non-conscious processes, then it seems that the value of this evidence is further undermined. One likely candidate for such an explanation is attention. That is because, attention, while directly linked to goal recognition in a manner that can easily map on to the hierarchical theory of intention that Pacherie has in mind, can also operate either consciously, or non-consciously. In a set of experiments, Kenridge et al. have shown how subjects are attending to primed stimuli, while at the same time not being consciously aware of the stimuli. More importantly, Kenridge et al.’s experiments set out to uncouple the commonly assumed connection between visual attention and conscious awareness. In the experiments, “spatial cues speeded detection or discrimination of targets subsequently appearing in the cued location, compared with those appearing elsewhere.” However, this spatial attention did not engender visual awareness, as the subjects consistently denied knowledge of the cues. They concluded that “spatial attention cannot be a sufficient condition for visual awareness and, moreover, visual attention and visual awareness cannot depend upon identical underlying neural processes.” This evidence is compelling, and suggests a further problem with the evidence from the Fourneret et al. experiments mentioned. Namely, the assumed conscious awareness of a disconnect is predicated on the

---

84 Kenridge, Nijboer and Heywood, “Attended but Unseen,” 865.
85 Ibid., 867.
assumption that conscious awareness of the visual cue (for the direction of the line) was constitutive of the attentional resources directed in order to effect the correction. However, Kentridge et al.’s evidence suggests that both attention and awareness could be working quite independently of each other, and, in relation to Fourneret et al.’s experiments, that there is no reason to think that anything about the awareness of the position of the hand (or, the sense of agency ascribed to the hand), was happening in virtue of an awareness about the visual cue. Returning to Pacherie’s original point about the role of consciousness in broader goal representation, what the evidence from Kentridge et al. suggests is that attention to the target can happen non-consciously, even when the stimuli in question is a visual input. If correct, this would directly undermine the premise of both the previous experiments. Again, more evidence is needed in order to hold that the more successful corrections were effected in virtue of a conscious awareness of this disconnect.

Moreover, we can wonder if there are ways in which attention can work non-consciously with intentions in order to affect agency. One alternative is offered in Wayne Wu’s useful characterization, where intentions are mental structural biases which elicit a given selection for the resolution of the many-many problem. The many-many problem is simply the problem offered up to an agent in the face of an array of possible targets of action within a behavioural space. In order to act, the agent must resolve the many-many problem, and Wu’s solution is that attention, guided by intention as a standing structural feature of our mind, is precisely what does this job for us. So, within the example of correcting the trajectory of a line as it is being drawn, it seems likely that the explanation

---

for the correction can be provided by an appeal to attention, with the presence of conscious awareness of the corrections not being necessarily relevant to whether they are carried out or not. For consciousness to perform a functional role in agency, in this case at least, it would have to be shown that the corrections depended on the fact of the conscious awareness itself. It seems, however, like this function can be carried out by attentional processes irrespective of the presence of consciousness. Moreover, as standing structural features of the mind, the link between consciousness and intentions, on Wu’s account, seems tenuous at best. We are therefore left with further questions about how to explain action guidance in virtue of conscious states.

3.4 Action Planning

Let me now turn to Pacherie’s account of what happens before action initiation. She focuses on Bratman’s theory of advance planning and the formation of prospective intentions. As Pacherie explains it, according to Bratman, intentions involved in advance planning have two essential features: they are both ‘terminators of practical reasoning’ and ‘prompters of practical reasoning’. The first refers to the fact that once a course of action has been determined, then no further reasoning is required to determine some new course of action about the matter in question. They are prompters of practical reasoning to the extent that, once a new course of action has been determined, practical reasoning about the means to achieve that action might be in order. While both features are important, for

---

Pacherie it is the second one that is relevant in order to explain a possible functional role for consciousness in agency. That is, while the formation of a prospective intention could well be the product of a conscious process of practical reasoning such as deliberation, the relevant issue is whether those intentions are related to “immediate intentions specific enough to ensure that the action is actually produced.” She focuses on two cognitive processes which can help in advance planning of this kind. The first is instrumental reasoning (IR) where, “we have instrumental beliefs about means-ends relations as well as various beliefs about how the world is and we use a relevant subset of these beliefs as premises in our reasoning from intended ends to intended means.” As such, IR is primarily concerned with the use of semantic information. Second, she describes the process of mental time travel (MTT) which is, “the faculty that allows a person to mentally project herself backward in time to relive past events or forwards to pre-live events.” This, in turn, makes MTT primarily concerned with episodic information. Moreover, these two concepts “should not be seen as mutually exclusive conceptions of what action planning is about, rather they may often function as complimentary processes.” Moreover, we might intuitively think that the processes described by IR or MTT are conscious events par excellence, and if Pacherie can indeed provide a link to how these

89 Ibid.
90 Ibid., 41.
91 Ibid.
92 Ibid.
processes are positively related to actions in the future, then it might seem as if we have found a clear role for conscious mental processes in action production.

The relevant issue is probing just what is the role of a process of practical reasoning in setting prospective intentions, and whether these intentions, once set, will prove efficacious in producing an action. However, considerations about automaticity undermine our confidence in the efficacy of such processes. Evidence has accumulated to the effect that agents will undertake actions regardless of their avowed beliefs or decisions. As we have seen, Tamar Gendler presents us with an array of such instances. Authors like B&M point to phenomena like aversive racism to note the strength that our implicit attitudes in the form of implicit bias, have on our behaviour. 93 And, as I mentioned, in Nomy Arpaly’s favourite example, Huck Finn deliberates extensively about whether to turn in his escaped slave friend, Jim. However, despite his conviction and ensuing decision that he should do so, he completely fails to carry out the deed. 94 These examples from Moral Psychology are an illustrative sample of the evidence that can be used to undermine our confidence in the efficacy of conscious deliberation. In view of this, it is not clear why we should be confident that undertaking processes like MTT or IR will result in our intended action. IR seems especially vulnerable to this type of concern. It seems to refer to specific, step by

step decisions about the undertaking of a future task. This is just the sort of deliberative activity that seems to be severely in doubt given the above considerations.

However, this may not be the case for MTT. Pacherie relates this distinction to that drawn by Gollwitzer\textsuperscript{95} between, “goal-intentions (‘I intend to do x’) and implementation-intentions (‘when situation x arises, I will initiate the goal-directed response y’).”\textsuperscript{96} If we agree with concerns about the efficacy of IR processes given the above considerations about automaticity and examples from Moral Psychology about belief-action discordance, then MTT and implementation intentions seem to be the final bastion of efficacious conscious mental states, on this account at least. As Pacherie points out, there is compelling evidence suggestive of the efficacy of implementation-intentions in action production. Strategies for ensuring the efficacy of implementation-intentions essentially involve co-opting features of automaticity not in the service of intention-formation (via conscious processes) but in terms of intention-acquisition (where the role of consciousness in the formation of an intention is irrelevant, even if one is conscious of the intention).\textsuperscript{97}

The final question is, then, whether the sort of episodic information handled by MTT needs to be conscious in order to be effective. Pacherie’s central evidence for this is the contention that MTT is a form of conscious thought. This is in view of neuroimaging evidence that seems to point to goal directed MTT’s engagement of working memory.\textsuperscript{98} And this is because, as Pacherie explains, “the engagement of working memory resources is generally considered as the hallmark of consciously controlled processes.”\textsuperscript{99} To be clear,

\begin{flushright}
\textsuperscript{95} Gollwitzer, “Implementation Intentions.”
\textsuperscript{96} Pacherie, “Conscious Agency,” 42.
\textsuperscript{97} Pacherie, “Conscious Agency,” 42.
\textsuperscript{98} Ibid., 41.
\textsuperscript{99} Ibid., 41.
\end{flushright}
in this case MTT would be a strategy for navigating the prevalence of automaticity in our behaviour, which makes the already limited role of consciousness in action production even more circumscribed to those times when this might be possible. Moreover, problems in the experimental design from the evidence cited by Pacherie seem to warrant further consideration. Namely, it is not clear why we should think that the engagement of working memory is necessarily linked with the use of MTT. In one of the experiments, “participants were asked to imagine themselves (i.e., simulate) actively solving a problem. Participants were presented with imaginary scenarios, in which they needed to solve several specific problems and were given cues that could aid formulating a plan to solve each problem.”

The large cognitive load implied by such an undertaking throws into doubt whether we should consider the engagement of working memory as a feature of MTT, or merely as a feature of taking part in this experiment.

Finally, let me mention a further tension with future planning within the account presented by Pacherie, this time as it relates to action guidance as defined by Frankfurt. On his account, agency is defined not by the causal antecedents of an action, but rather, by the guidance mechanisms of the agent vis à vis the action as it unfolds. If Pacherie wants to subscribe to Frankfurt’s view, then this seems to further undermine the role of MTT and IR. While it may be true that they have a role to play in the production of action, on this definition it does seem like this is the same as saying that they have a role to play in agency. In that case, while conscious processes like MTT and IR may have a facilitating role to play in action, it remains to be seen whether any conscious processes are involved in agency.

---

itself. Therefore, we are left with a severely limited role for consciousness in action production, and one which is not related to agency as originally defined.

Features of action must happen in virtue of a conscious state’s being conscious in order for us to think that consciousness does have a functional role in agency. Contra Pacherie’s view, I have argued that there is no place for a functional role for consciousness in agency on her account. On the available evidence, conscious mental states seem to merely be present in tandem with agency as it unfolds. Their mere presence is no proof, however, of their functional efficacy. Likewise, the efficacy of prospective intentions is doubtful, and experimental data in support of the role of consciousness in MTT is questionable. Moreover, if Pacherie agrees with Frankfurt that action is really about action guidance, irrespective of causal antecedents, then it is not clear how on her view, what might count as a limited role for consciousness in action production via MTT and IR, gets to count as a functional role for consciousness in agency itself. What should be clear from these considerations is just how far our folk theoretical assumptions about agency seem to be from the picture painted by science. And, it is in light of this that we should examine our account of human moral psychology, and be open to the idea that automaticity is pervasively involved in it.
4 Chapter: Automaticity in Moral Psychology

Bearing in mind the possibility of intelligent automaticity, let me examine now some of what has been taken to be at the heart of automatic action by some moral psychologists. First I will examine Tamar Gendler’s account of alief which will aid in our discussion of the normative status of automatic behaviour at large. While she is mainly concerned with actions which are not in accordance with our avowed beliefs, the second section will examine the view defended by Brownstein and Madva, which takes up Gendler’s account of alief, but instead focuses on automatic actions which can be said to have self-standing normatively assessable status, irrespective of the role of belief. This will allow me to contrast their account of what gets to be called a moral action, and examine an account of automatic behaviour which is normatively assessable, but with which I will nevertheless disagree. Brownstein and Madva will in the end argue that automatic behaviour is assessable in virtue of being pro tanto norms responsive. I will end the chapter by questioning whether this is a satisfactory account of moral assessibility.

4.1 Gendler’s Aliefs

Gendler sets out to establish a case for the discordance between rational belief and the actions and behaviours of individuals. That is to say, between their conscious, occurrent beliefs about any given subject, and the discordance of their actions with the content of that
belief. Gendler points out that there are mechanisms at work in the background that are eliciting behaviours other than the ones concordant with the conscious or avowed beliefs of a given individual. On her account, the difference between someone feeling secure while walking onto a solid wooden platform, versus someone feeling reluctant about walking onto a (guaranteed to be safe) transparent bridge over a chasm, is that there is a mismatch between the visual inputs and the consciously held belief. She attributes these instances to what she dubs the “belief-discordant alief.” The content of these is roughly whatever is evinced by the behaviour – so, if I am reluctant to eat vomit shaped plastic that I know is sterile, there is an alief at work that roughly says “Vomit! Dirty! don’t put in your mouth!” even if I have a concurrent, occurrent belief to the contrary. And so on for all like cases.

We might wonder why one should think about an alief at all, or how it is that Gendler arrived at this term. On her account, an alief is:

“… so called because alief is associative, automatic, and arational. As a class, aliefs are states that we share with nonhuman animals; they are developmentally and conceptually antecedent to other cognitive attitudes that the creature may go on to develop. And they are typically also affect-laden and action generating.”

Gendler argues that a notion like alief is necessary to explain the various phenomena she points to. However, as we have seen, there is already a broad literature for something called automaticity, which captures the wide range of phenomena she mentions. Nevertheless, for the remainder of the chapter the discussion will centre around aliefs, given that it is the direct point of contact between Gendler and B&M.

On Gendler’s account, alief might arguably be an important epistemic category in order to ascribe different contentful mental states as the causal sources of agentive

---

103 Ibid.
behaviours. And yet, as we have seen, it is not the case that we can say that for instance *all* automatic behaviours are the product of cognitively penetrated skilled action. As per the view endorsed by Fridland, automatic behaviours exhibit a variety of features, none of which are the exhaustive, or necessary list of any one behaviour. Some behaviours might include some of the features, some not. More importantly, the necessity and sufficiency conditions for all of these features has been empirically undermined, and not only do none of these features appear to be sufficient or necessary, they are also not necessarily present together.¹⁰⁴

In the above characterization then, we see indications of the sort of intelligent versus unintelligent dichotomizing to which philosophers and psychologists are prone, according to Fridland. It is for example, not clear why we should qualify aliefs as arational. Clearly if they are tied to contentful states, even if they are not in accordance with occurrent conscious contentful states, then they are behaviours that are happening in virtue of reasons carried by those other contentful states. They are not arational, they are simply not directly caused by a deliberate, conscious intention.

It is also important to note that it is not Gendler’s claim, on any of her examples, that there is a feeling of alienation from her limbs, or from the subjects and the behaviours that they undertake. It is a different situation if there is a patient that suffers from anarchic hand syndrome, in which case the patient’s limbs undertake behaviours of their own accord, also in discordance with the conscious intention of the individual. What is important about these automatic cases, different from the anarchic hand cases, is that there

---

¹⁰⁴ For a thorough account of all the empirical evidence undermining these features being necessary or sufficient conditions of automaticity see Moors and De Houwer, “Automaticity, A Theoretical and Conceptual Analysis.”
is no comparable disruption to the sense of agency while we undertake the belief discordant action. That is, we experience ourselves as undertaking an action, and even know why we are undertaking an action – what catches us with surprise is how this action lags or is in conflict with the conscious thoughts that we are currently entertaining at the time. It is therefore important to keep in mind that we are not here talking about instances of alienated agency. With that in mind, Gendler’s provisional definition of an alief is the following:

A paradigmatic alief is a mental state with associatively linked content that is representational, affective and behavioral, and that is activated – consciously or nonconsciously – by features of the subject’s internal or ambient environment. Aliefs may be either occurrent or dispositional.¹⁰⁵

Her definition is sufficiently broad to apply to brain states which may or may not be linked to representations, the nature of which remains an open question as well. What is of central interest is what she takes to be the philosophical implications of aliefs to be, specifically for moral psychology. What is in question is the role of beliefs, and especially the role of conscious, occurrent beliefs. More importantly, whether our actions have to be in accordance with those beliefs in order for our actions to remain morally assessable.

4.2 Moral-Worth and Belief

Let me then move on to what this means in terms of an action’s moral standing. Gendler’s conclusion is that in light of actions’ susceptibility to aliefs, “in order to live well, we must work to bring our habits in accord with our reflective beliefs.”¹⁰⁶ Gendler refers to aliefs as “recalcitrant norm-discordant responses.”¹⁰⁷ On her account, times when belief-discordant

¹⁰⁶ Ibid., 662.
¹⁰⁷ Ibid., 553.
aliefs are useful are limited to those life enriching moments where associative chains are usefully exploited for enjoyment, such as in the thrill of bunjee-jumping, or through the different ways in which activating aliefs through our imagination via different art forms would be beneficial. However, in general, it is what she calls a teleofunction-concordant alief that is by and large conducive to our wellbeing and self interest (hence teleofunction), and belief-concordant aliefs are the ones that are responsive to the right norms.

With this, Gendler moves into her prescription for what to do with discordant aliefs – and taking a cue from Plato, she alludes to the ancient ideal of a well-regulated soul being analogous to a soul that has harmonized all its parts. Recall that in the Platonic picture, the soul is made up of several parts, and the intellect is tasked with regulating the lower parts in favour of aligning the entirety of the soul with its highest feature. Gendler highlights two strains of the tradition that are usually appealed to in order to address disharmony of action and intention. One of those is through habituation and rehearsal, more in line with the Classical Greek tradition, and the second is “through the refocusing of attention (especially by directed imagination), thereby redrawing the lines of internal association.”108 In the end, both strategies are geared toward bringing our belief-discordant aliefs in line with our considered beliefs on a given subject matter. Gendler then moves to consider the implications of alief for cases such as those of aversive racism, where implicit attitudes, or what has been popularized in the media as implicit bias, plays a crucial role in the often-tragic outcomes of racialized minorities as they are victimized by the authorities, or simply by how they are treated in their day to day by average citizens. What is crucial, of course, is that in all these instances, the oppressors would be avowed non-racists. The relevant

108 Ibid., 572.
question is, then, how one might correct one’s behaviour once it is revealed that one is committing an act that is discordant with one’s own moral commitments. In these cases it seems clear that the natural goal would be to want to bring one’s behavioural dispositions in line with the norms one has chosen to avow as one’s own morals.

However, while it may be true that in cases like those of aversive racism we need to bring our actions in line with our avowed beliefs, B&M also point out that there are categories of automatic action which cannot be subject to this requirement. While they take the general category of alief as a possibly helpful one to explicate what is causally at play with automatic action, they posit that the fundamental question is “not simply how to regulate, control, or change automatic dispositions but rather what automatic dispositions are (and are not) good for and when they can (and cannot) be trusted.” On their view, Gendler overstates the way in which aliefs are impervious to norms in the environment. They offer an account where automatic behaviours do in fact express a sensitivity to norms, as the agent moves to alleviate ‘felt tensions’ in their environment, “everyday examples include moving around until one has found the right spot for viewing a painting or stepping backward in order to alleviate the subtle discomfort induced by a ‘close-talker’.” On B&M’s account, there is an entire category of automatic behaviours that are assessable in their own right. I have already pointed to two of their central examples in the introduction. One is Barack Obama’s handling of Justice John Robert’s misreading of lines in the swearing in ceremony at his inauguration. Brownstein and Madva argue that it was in virtue of Obama’s socially atuned senses that he was able to respond appropriately and save the

---

110 Ibid., “Ethical Automaticity,” 69.
situation from further embarrassment, by smiling just right. What is praiseworthy is the level of social skill attributable to Obama in the handling of this situation. An even more dramatic example, meant to address concerns about whether the behaviour was merely prudential, instead of genuinely moral, is the case I have already recounted of Wesley Autrey. This individual was at a subway station with his two daughters when he noticed an epileptic fall into the tracks because of a seizure. Without hesitation Autrey jumped into the tracks just in time to hold the man down while the train passed overhead, saving his life.\textsuperscript{112} It would seem that this kind of spontaneous heroic action would insulate their view from any concerns that alief as they construe them are not normatively interesting.

B&M’s account centers on the role of reactive attitudes.\textsuperscript{113} Importantly, these express normative relations of approbation or disapprobation within our social relations, such that as one responds to the norms in these relations they are a response to a sense of oughtness in a situation. One of their central examples is one that is originally from Merleau-Ponty,\textsuperscript{114} where a subject perceiving a painting usually does their best to place themselves in the optimal place in relation to the painting such that they might perceive the painting at its best. On Merleau-Ponty’s account, this generalizes to all subject-object interactions. By and large, the agent’s perception of an object’s sense/signification is mutually constituted through the agent’s ongoing navigation of their milieu. One can see from the example that Merleau-Ponty does not only have mere objects such as cups in mind when he thinks about this, but also value laden perception, such as the agent’s beholding

\textsuperscript{112} Ibid., 84-85.
\textsuperscript{113} Ibid., 77.
of a painting as they seek to achieve the optimal aesthetic vantage point. Regardless of the accuracy, or theoretical baggage of such a characterization, it is, I think, at least intuitively appealing in that we can all relate to the experience. On B&M’s account, the adjustments happen in virtue of ‘felt tensions’. Against alief, they argue that the re-adjustment one accomplishes automatically in bodily orientation towards the painting is not in virtue of a contentful state like “Really big painting! Disorienting! Move away!” but that, subject to a multiplicity of motor adjustments, one single alleviating state is unable to account for “the way the affective and behavioural components unfold and interact over time.” To put it another way, it is not the case that the stimuli inputs of the painting can activate a representation of what is the right or wrong way to view the painting, because the rightness or wrongness of viewing that painting is being constituted as it unfolds.

They take such cases where there is norms responsive behaviour that nevertheless could not be in concordance with any kind of belief to be evidence against the TDH thesis. Specifically, while it may be true that a general belief about the nature of picture viewing might be associated with the behaviours, the argument is that the motor corrections are so small and contextual that to explain action in virtue of behaviour responding to that belief would not offer any explanatory value, and instead makes sense as responding to sociocultural norms, which in this case would presumably be related to aesthetic preferences. On their account, automatic behaviour is responding to an ‘F-T-B-A’ relation; that is feature-tension-behaviour-alleviation. Features refers to the sort of

---

116 Ibid.
117 Ibid., 420.
"environmental properties that make certain possibilities for action attractive."

Tensions refer to the sort of push and pull of valences which are either negative or positive in relation to a given experience. These tensions are attended to at the level of the agent’s phenomenal awareness. The behaviour of the agent is therefore implicated both in the ongoing undertaking of a behavioural space such that these tensions become manifest, and then an ensuing shift toward alleviating those tensions as they arise. In this sense, while on the view they attribute to Gendler the agent should strive for a harmony with their avowed beliefs in their behaviour, in B&M’s view the agent does strive for a harmony with their environment. Given their account of tension alleviation over time, aliefs are importantly norms responsive. In fact, it is in virtue of being norms responsive that aliefs are efficacious.

One of the crucial requirements for showing that aliefs are norms responsive is showing that they are susceptible to self-modification. Given the necessarily changing nature of behaviour, specifically at the micro-level B&M have in mind, aliefs are constantly being modified as the orientation of the agent changes within their environment. In this way, “within immediate contexts, activated aliefs modify themselves by eliminating themselves. Rather than by updating to reflect the cumulative evidence [like a belief], aliefs self-modify by compelling the agent to change her bodily orientation to the world so that the source of tension vanishes.” And what is true at the micro scale, is true at the macro scale as well. As the accumulated experience accrues overtime, the nature of the aliefs that

118 Ibid.
119 Ibid., 424.
are there for an agent to have triggered is also modified gradually. Merleau-Ponty refers to this as sedimentation.

A further feature of being appropriately norms-responsive is susceptibility to error. While on Gendler’s account, an alief’s discordance with a belief makes its susceptibility to error obvious, for B&M, “the museumgoer, perhaps as soon as she steps back, she comes to feel too far from the painting. It could be that the appropriate behaviour was to squint her eyes and tilt her head, rather than to move away.”120 They provide a further case where “an interlocutor leans in to whisper something important. The addressee might rightly perceive the inward lean but under – or overreact affectively, by being coldly indifferent or disconcertingly solicitous.”121 In this way, while the valence might be rightly perceived, a different feature of the alief’s F-T-B-A structure might misfire.

Much like Fridland in the first chapter, Brownstein and Madva also see alief, and certain automatic behaviours as falling somewhere between intelligent and unintelligent. But, they stress that this tilt towards possible intelligence should not be taken too far. Specifically, they want to argue that while aliefs are norms-responsive, they are not reasons-responsive. On their view, even if there is per force an explanatory reason for why an action is undertaken, this does not mean that an action is taken in virtue of that reason. As they explain,

> for a reason to explain an action-guiding state in non-deliberative contexts, the reason must be at least available for practical reasoning. On a view we find intuitive, practical reasoning is the capacity for resolving, through reflection, questions about what to do… An action-guiding state is reason-

120 Ibid., 424.
121 Ibid., 425.
responsive just insofar as it is capable, *ceteris paribus*, of figuring in practical reasoning in the right ways, perhaps by mediating inferences about the appropriate means for achieving one’s ends.122

From their characterization of alief, we can see how this conclusion is likely if, essentially, these sensitivities to valenced tensions are too minute and quick to be available for anything like being formulated in the service of practical reasoning. In the end B&M argue that alief provides a sense of oughness to the actions which it elicits. However, this oughness can be defeated since, given that aliefs are pervasively involved in action production, they are also responsible for the actions undertaken by an aversive racist. Therefore, aliefs provide pro tanto normative justification, unless, ironically, the action they elicit turns out to be immoral, which seems to beg the question. It is like saying an alief has pro tanto normative justification just in case it does, in which case it is not clear what the source of this justification would be in the first place. B&M respond by first saying that in this regard alief fares no worse than belief itself, and that, as it turns out there may be cases where alief is actually normatively superior than belief. Here they point to the example I have already alluded to, from Nomy Arpaly. In this case, instead of an aversive racist, we have in Huck Finn an “aversive egalitarian.”123 Huck is convinced that he must turn in the escaped slave Jim, and his reasoning is informed by his racist belief system. However, as we know, Huck fails to turn Jim in despite ample opportunity. In this case the alief guided behaviour that made Huck act as he did proved normatively superior to the beliefs he had on the matter. However, if all B&M have to ground the normative value of Huck’s behaviour is this version of aliefs, it would seem Huck might fare better with an account

122 Ibid., 428.
123 Ibid., 429.
that can provide him with substantive reasons for action, even if he is believing the wrong thing. This is the topic of the next chapter.
Chapter: Automaticity and Acting for Reasons

In Chapter 4 I discussed how Tamar Gendler’s view sets the bar for the assessability of an action at the level of its concordance with our endorsed beliefs. However, through Brownstein and Madva I made the case that there is at least a full category of actions which could not be subject to belief in order for them to be morally assessable along the lines of what Gendler has in mind. On B&M’s account, automatic actions are norms responsive in virtue of their attunement to their environmental milieu, and this is enough to view them as morally assessable. This is an attractive view, but we might wonder if it falls somewhat short of our common understanding of moral agency. After all, it might be that the actions themselves are norms responsive and assassable, but this might not mean that the agent herself is. We might think that it would be better if we could account for the agent’s actions along moral reasons. If we consider the weak position of the account of pro tanto norms-responsiveness that I alluded to in the previous chapter, we might think that an account of acting for reasons can provide us with more substantive normative grounds on which to ground the status of automatic behaviour. Therefore, let me first examine what an account of acting for reasons in virtue of automatic processes would look like.

5.1 Deliberation

For A&S, to think that actions depend upon a further action of deliberation can only set us off on a vicious regress. To be clear, they are specifically referring to deliberation as a mental act which is done consciously, in favour of attempting, through conscious thought,
to figure out what to believe or do. So, while this may not capture everything one might take deliberation to be in the broadest sense (perhaps there is non-conscious deliberation), it nevertheless captures what we would take to be the sort of conscious reflection related to the avowal of a belief, especially, the avowal of an occurrent belief along what Gendler has in mind. For A&S, this is simply not how humans function. Instead, we are primarily driven by non-deliberative processes (ND), which can, to be sure, be supplemented by deliberation, but which can, and do, work entirely independently of deliberation. This is not to say that ND processes are necessarily conscious or non-conscious. But it is to say the following. First, that deliberation is only possible because of ND processes themselves. And, that deliberation can be non-efficacious in the production of action even after it is engaged by ND processes. In the end, it is only ND processes which bear causal responsibility for action production, and acting for reasons.

A&S endorse a desire based view of Moral Psychology. However, they still want their moral agents to be acting on the right reasons, the same way a non-sentimentalist would like their agent to be. If we accept that acting for reasons is grounded upon acts of deliberation, then it would seem like an account of morality which is based on desires might be in trouble. That is because desires are often associated with appetitive, impulse driven behaviours that are not subject to prior reflection. As they note, deliberation plays a central role in moral theorizing. We deliberate about whether to give to the poor, and if so, how much. We deliberate about what kind of a person we would like to be. We deliberate about the right approach to take as a human in relation to the wider planetary ecosystem.

125 Thank you to Myrto Mylopoulos for pressing this point.
Deliberation seems central to acting for reasons. Indeed, as we saw, it is precisely this inability of aliefs to be available for deliberation which made them, for B&M, unable to count as acting for reasons. However, on A&S’s view, it is not in virtue of a deliberative act that humans act for reasons. If this were so, then acts of deliberation about reasons would be subject to a vicious regress as we attempted to deliberate about deliberating, about deliberating, etc. *ad infinitum.*

Deliberation can be either practical or theoretical. It can be about figuring out what to do, or about figuring out what to believe.126 Moreover, on A&S’s view, deliberation is an action. For the skeptic about whether deliberation is an action, A&S point out that it bears all the phenomenal traits of an action: it is a behaviour that can be seemingly carried out at will (it carries with it a sense of voluntary agency); it can also fail like an action would – one can fail to deliberate when one would want to deliberate; moreover, deliberation also falls prey to the sort of “aberration, distortion, and perversion characteristic of overt action.”127 Here A&S have in mind times when one might deliberate in an absent minded way, deliberate and feel alienated from the very deliberation, or fail to stop deliberating despite wanting to, among others. In the end, deliberation is an array of mental actions geared towards achieving a specific end. Most relevant for our purposes, the sort of deliberation A&S have in mind is constituted by conscious mental states. As they explain, “for a mental activity to be deliberation, it must also be carried out by means of particular kinds of mental acts: bring to mind –to consciousness—various ideas (general or particular, abstract or concrete) or images (visual, auditory, gustatory…).”128

---

126 Ibid., 22.
127 Ibid., 23.
128 Ibid., 24.
Additionally, it is not sufficient that they be simply brought to mind. They must be brought to mind for a specific end, and they must bear the right kind of “rational relation to the propositions one considers, with the aim of reaching a conclusion.” They nuance their characterization of deliberation further, however, by pointing out that this does not imply that deliberation proceeds in the kind of neat orderly way one would like to imagine, and it highlights the role of automaticity even in these actions which are thought to be voluntary and controlled par excellence. They see deliberations as made up of small units of activity thrown up into consciousness like fragments directed towards a specified end. These fragments bear the right kind of rational relation, but they happen in a messy sort of process. Another important point is that a process of deliberation is not guaranteed to be reasonable; one may undertake an act of deliberation for the wrong reasons, or at the wrong time for one to be doing so.

When it comes to thinking and acting for reasons – if we agree that deliberation is a kind of action, then we can see how the regress objection will unfold. Namely, if we need an act of deliberation to act for reasons, then a further act of deliberation in order to deliberate is needed – as I mentioned above. A&S consider the possibility of “Previous Deliberation”, “Present Deliberation”, and “Possible Deliberation.” However, they all run into the issue of a vicious regress when it comes to having a kick starting reason to begin the process of deliberation. Previous Deliberation seems to be the likeliest contender for thinking that actions are being done in virtue of reasons due to a deliberative act. But, for a previous deliberation to have been undertaken, it would have been an action that had

129 Ibid.
130 Ibid., 25.
131 Ibid., 29-33.
happened for a reason; a reason which, if we stick to our idea about deliberation, would have had to have been determined by a previous act of deliberation. A&S conclude, then, that acting for reasons cannot be done in virtue of acts of deliberation, and that it must instead happen in virtue of non-deliberative (ND) processes. These are automatic processes, to use the terms I have been using so far. So, on their view, acting for reasons is underwritten by automatic processes. We can see then that, to an extent, the taking up of reasons for action proceeds automatically. That is to say, the selection of a reason for action is not subject to an intention about selecting that reason for action. The reason is simply selected automatically via ND processes.

5.2 Acting For Reasons

The question is, then, what an account of acting for reasons through automatic processes might look like. A&S develop a compelling account. They first draw a distinction between objective reasons and rationalizing reasons. Objective reasons are external reasons; that is, reasons external to the agent’s own mind. However, acting and thinking for reasons does not simply depend on objective reasons. Acting and thinking for reasons also depends on internal reasons, that is – reasons related to the mental states or beliefs of the agent.

An objective reason is grounded by the facts of the world. It is or is not the case that my coffee has had sugar added to it. However, I might think there is no sugar in it, but unbeknownst to me, my partner kindly added it for me before bringing it to the table. In this way I may have a good internal reason to add sugar, but vis a vis the facts of the world, I do not. Internal reasons for which we act may or may not cohere with external reasons
for which we might act. Acting for a good or bad reason does not depend on the coherence of the internal and the external reason. Instead, it depends simply on the rational justification of the internal reason. That is, I may be quite objectively mistaken about the facts of the matter, but all things being equal, as far as I might know in that situation given my present access to knowledge, I may be acting for quite good, although objectively incorrect, reasons. It is in virtue of the logical relation between the content of my rationalizing reasons that I am acting for good reasons. As they explain:

*To think or act for a reason* is for the event of one’s thinking or acting to be caused (or appropriately causally explained) by one’s other attitudes in virtue of the fact that these attitudes rationalize (to some extent) the thought or action.132

The metaphysics of causation in virtue of properties is complicated and best left to other theorists of the pertinent subjects. It is enough for A&S to subscribe to the weakest possible thesis by which some form of causation in virtue of contents or the contents of properties is possible, given common sense avowal of such possibilities. Moreover, the causal relations are predicated upon the right logical or probabilistic relations of causation. On this account, acts of deliberation are useful in helping us survey the logical relations at play, but not in explaining the causality in virtue of the logical relations. The causally

132 Ibid., 62.
efficacious mental states arise in virtue of ND processes, and they come into play due to the logical relations that bind them together.

5.3 Desire

Finally, it is worthwhile discussing A&S’s account of what they call intrinsic desire. The reward system of the brain is the only system which we know the cause of all the normal effects of intrinsic desires; it is a cause of “overt and covert action, of positive and negative feelings, and of the cognitive effects most associated with intrinsic desires.”\(^{133}\)

A&S’s theory of desire is as follows:

Reward Theory of Desire: to have an intrinsic desire regarding it being the case that P is to constitute P as a reward or a punishment.
To have an intrinsic appetitive desire that P is to constitute P as a reward.
To have an intrinsic aversion to P is to constitute P as a punishment.\(^{134}\)

The important point here is that desires are constituted by the reward system of the brain. Just like rationalizing reasons and objective reasons need not be one and the same, neither does the appraisal of reward or punishment need to be tied to an objective fact of the matter about whether something is good or bad. As they explain, “what makes a conceivable state of affairs P count as a reward or punishment for a given person… is constituted by the contingent power of that person’s representation that P to cause… a reward or punishment signal.”\(^{135}\) Part of the content of these representations seems to be something innate, and

\(^{133}\) Ibid., 127.
\(^{134}\) Ibid., 127-128.
\(^{135}\) Ibid., 132.
some are constituted as either reward or punishment signals as we get older. What this suggests is that the entire example suite of A&S is made up of events which are too complex to be constituted as a single reward or punishment. Instead, we can arguably construe it in terms of micro-rewards and punishments to myriad valences that are being activated, much like the Brownstein and Madva view. It is not just that giving to charities constitutes a reward, it is that it is part of a reward system from which the desirability of giving to charities can be derived as a reward if it were to arise as an event where an action needed to be undertaken. So, the person may never have had a chance to give to charity, but their suite of intrinsic desires will establish the sort of signal something like this new experience will give rise to, plus a new bit of information about that new instance. A&S describe their ideal picture:

Imagine an organism constantly perceiving and cognizing its environment. Through a combination of genes and environment, some of the states of affairs it perceives and conceives are constituted by one of its learning systems as greater or lesser rewards, while others are treated as neutral. The net sum of these rewards comes to some amount of net reward at each moment. At the same time, the organism also unconsciously, perhaps associatively, predicts a given amount of the reward for that moment. The net reward of the moment has subtracted from it the predicted amount of net reward, and the result is either a positive value, zero, or a negative value.  

5.4 Assessability

A&S point out that “there are two kinds of moral responsibility of which one might give a theory: accountability and attributability.” Accountability refers to the manner in which someone should be treated, should we decide that they are responsible for an action. Attributability is related to the sort of attitude we might hold towards a person if we impute

\[136\] Ibid., 135.

\[137\] Ibid., 160.
them with a given action. A&S’s view is then an account of “attributive praise- and blameworthiness.” On their account, what is praise or blameworthy is whether a person has good, or ill will respectively. Below is the definition for good will as they understand it:

Complete good will is an intrinsic desire for the right or good, correctly conceptualized (i.e., presented via the correct sense, mode of presentation, narrow content, primary intension…)
Partial good will is an intrinsic desire for some part of the right or good, correctly conceptualized. A person has more good will insofar as her intrinsic desire for the right or good is stronger, and less good will insofar as this intrinsic desire is weaker.

Ill will is just as per the above, substituting ill will in place of good. In addition, a person can be morally indifferent – when they lack good will; and they can be reverse morally indifferent – when they lack ill will.

Associated with good will are the suite of intrinsic desires and their familiar cognitive effects such that a person desires to do good, and experiences the attendant feelings and thoughts of such a desire. The reverse would be true for ill will. Importantly, these are the effects usually associated with these desires, but are by no means necessary – nor do they constitute an impulse to act. Instead, the desires are “the grounds of our abilities to act for reasons… They might only contingently move us to act, but they are what makes us poised to be agents in the first place.” The referent for the intrinsic desires associated with good will are determined by the right normative theory about the nature of the good. In that way a deontic reading of the Good would result in the attendant concepts as being

---

138 Ibid.
139 Ibid. 162.
140 Ibid., 163.
the right concepts associated with good will. This would likewise be true of a consequentialist.

On A&S’s view, most people have partial good or ill will. On their view, “partial good will can be thought of as an intrinsic desire for something that there are pro tanto moral reasons to bring about, while partial ill will can be thought of as an intrinsic desire for something that there are pro tanto moral reasons to not bring about.”141 On their theory of praise- and blameworthiness, the agent is worthy of those attitudes in virtue of the good or ill will that was part of her rationalizing reasons. On their theory of virtue then, a person is virtuous or vicious in relation to the degree of good or ill will that they have, as well as their lack of ill will and degree of moral indifference. And, more specifically, “a virtue is an intrinsic desire that constitutes either complete or partial good will, or perhaps a collection of such intrinsic desires…”142

The question for me is how I can link up what I have discussed about automaticity thus far with the theses put forth by A&S. The crucial point to be made is that on their account the cognitive life of a virtuous person will be noticeably different. It is worth quoting at length how this would be so:

Consider the kind person. She is more inclined to prevent suffering than most of us, and that is an obvious manifestation of her desire that people not suffer. She is also more inclined to feel bad about the suffering of others, and that too is an obvious manifestation of her good will. But the kind person is also going to have thoughts that are different from other’s thoughts. The kind person will be more likely to notice suffering: it captures her attention. The kind person will be more likely to remember that the person she met three months ago was struggling to come to terms with her brother’s death: it sticks in her memory. The kind person is more likely than others to see an angry man’s outburst as a product of his feeling threatened or humiliated: it seems plausible to her. And in many other

141 Ibid., 166.
142 Ibid., 202.
ways, the kind person’s cognitive life will be different from the cognitive life of the less-kind person.\textsuperscript{143}

On A&S’s account, it is in virtue of the contents of desires that “involuntary shifts in attention”; “changing dispositions to learn and recall”; “changes in subjective confidence”; and, “distortion by emotions and wishes” all play a role in shaping the cognitive life of an individual.\textsuperscript{144} Attention is affected by desires simply in virtue of the preferences people have to be attentive to stimuli they already have an affinity towards. This in turn leads to the formation of beliefs in virtue of having attended to new stimuli which presented new information; information which, without having attended to it, would not have been available for belief. A further point here is that these are instances where conscious awareness and attention go hand in hand. But, if we remember back to Kenridge et al.’s experiment in Chapter 3, conscious awareness and attention are not necessarily linked. And, what is more, neither is attention to visual cues necessarily linked to conscious awareness. In view of this, we can speculate about the scope of attention in gathering information, and the cast of such information gathering that would be available for a person of good will. Arguably, non-conscious attention would be constantly directed to relevant information, and this relevance would be shaped in virtue of the contents of the agent’s intrinsic desires.

\textsuperscript{143} Ibid., 255.
\textsuperscript{144} Ibid., 227.
Chapter: Conclusion

Mark Twain proves himself an uncanny observer of the unfolding of human morality and agency at large. The purposive, conscious construal of our actions can be plain and simple or elaborate and complicated. We can deliberate about what we should believe, deliberate about what we should do, and we can create plans based on these deliberations – the creation of a plan being a form of deliberation itself. And yet, it will not be in virtue of this sort of conscious deliberation, and the ensuing conscious intention, that we will act for reasons. Instead, as we saw from A&S, we will act for reasons based on a constellation of factors that are associatively linked in our background mental processes in virtue of the right logical and probabilistic links between them. And, it is in virtue of these factors that acts of conscious deliberation themselves will be undertaken. Acts of deliberation themselves will be useful for bringing out and surveying the associations and interactions, and, in other words, making explicit the associative links through which an action has been, is being, or can be undertaken. But, the mechanisms in virtue of which the actions themselves have been undertaken, have proceeded automatically. That is, they have proceeded without appeal to a conscious decision or intention, even if they are proceeding consciously.

If correct, what this picture makes clear is that our view of moral psychology is limited. Our folk theoretical account seems to tell us that first, consciousness is somehow constitutive of agency; and second, that our grounds for assessing rational behaviour is in terms of acting for reasons, where acting for reasons is underpinned by acts of deliberation. However, if on the one hand it is not clear whether anything about action is happening in
virtue of conscious mental states, and if, on the other hand, deliberation is not necessary in acting for reasons, then our folk accounts of agency and rationality are in trouble. Moreover, as we saw in Chapter 4, whereas our common FP assumptions might lead us to hold something like the TDH, automatic behaviour has at least features in virtue of which it can be pro tanto morally assessable. All these three sources of resistance for the assessibility of automatic behaviour as moral have therefore been undermined.

Of course, an accurate descriptive account (or, as accurate as possible) of agency and acting for reasons is necessary for a good account of human moral psychology. It has been
my contention that the right place to look is in accounts of automaticity. Here, then, is a schematic look at how my argument has proceeded:

1. Folk Psychology (FP), and by extension Moral Psychology, cashes out morally assessable action as action that (i) is planned, initiated, or guided by conscious states, and (ii) is done for reasons in virtue of being related to acts of deliberation.
2. We have reasons to doubt evidence purporting to show that actions are planned, initiated, or guided in virtue of consciousness.
3. We have reasons to doubt that deliberation is necessary for acting for reasons.
4. We need to re-assess our FP account of morally assessable agency.
5. Automatic processes are the primary way through which action is produced.
   a. Automatic action is pervasive in the form of action that does not occur in virtue of a conscious intention.
   b. Different automatic processes can proceed with or without consciousness.
6. Automatic processes are intelligent.
   a. One acts for reasons in virtue of automatic processes.
7. Moral Psychology: Humans act for reasons in virtue of automatic processes which can proceed either consciously or non consciously; the moral assessability of an action is tied to the intelligence encoded in automatic behaviour, which can be cashed out in terms of skilled behaviour.

Let me then provide some discussion of the above, with a view to summarizing the project at large, and in order to gesture at what I take to be interesting ways forward. Recall from Chapter 2, that I supported Fridland’s hybrid view of automaticity. Two features I stressed are both that automatic actions are actions that are not necessarily happening in virtue of
conscious states, nor are they actions that are necessarily guided by a specific intention. From this, given the considerations of Chapters 3, 4, and 5, we can glean that the scope of automaticity is pervasive, and morally assessable.

Crucially, as we have seen from Fridland, the crux of the issue seems to also lie in the dichotomizing of intelligent vs. unintelligent, and of equating it with conscious/controlled vs. non-conscious/automatic. As we have seen, the picture is much more nuanced and complex. One central aspect of this is that just because something is automatic it does not mean it is not conscious, or controlled. Likewise, just because something is non-conscious does not mean that it is wanton, or unintelligent, or uncontrolled. More nuanced views mean that our neat folk psychological picture of moral action as anchored in intention and deliberation; that is, as a conscious consideration of relevant information in the service of arriving at a suitable conclusion; needs to be revised in favour of including automatic processes in place of the assumed purposeful decision making that is central to many moral theories. The role of consciousness in action production and control remains unexplained – and while there is no doubt a role to be played for consciousness in agency, the fact that we are still unclear as to what it actually does, and that we have evidence that action can very well proceed without it, is enough to dispel our original ideas about what the relationship between consciousness and agency is all about. A similar story can, and has been, told about rationality and deliberation.

The example that has run through the thesis has been that of Huck’s own struggle with himself as he wrestles with the principles of goodness he has been brought up with. His is a case of ongoing deliberation with his actions actually being mediated by background automatic processes. In his case, we can assume there is at least one belief
about right and wrong operative in his reasoning. Namely, it is wrong to steal. And we can assume there is at least one prominent belief about the facts of the world also guiding his reasoning, (I express it in the form of what can be assumed to be the token of a more general belief in order to avoid very charged language) – Jim is someone’s property.

In what is at least one measurable bit of moral progress in the last 150 years, we no longer think it is tenable to hold that a human person is another human person’s property. In fact, we think this is a clear-cut case of immoral behaviour. In Huck’s case, his deliberative act in virtue of his avowed principles yielded the immoral conclusion that it was right for him to turn Jim in. However, Huck manages to act morally despite his wrongful deliberation. This suggests that he was acting for reasons other than those picked out by his deliberations; reasons which were not available to his conscious mental life, and were in fact at total odds with it. In fact, what is heartbreaking about Huck is just how very good he really is, and how ridden with self-loathing he is regardless. This further suggests that reasons for action are not necessarily tied to belief, and that the assessability of actions is not necessarily tied to belief. In fact, actions are assessable as praise or blame worthy on different grounds; that is, in virtue of the automatic processes through which they are undertaken. And that is because it is automatic processes, in the way of the Non-Deliberative processes that A&S have in mind, that are responsible for acting for reasons. And, those reasons may be praiseworthy, but at odds with our beliefs.

This is not to say that automatic processes that pick out reasons for action can somehow reveal the normative substrate of the world which should be the real guide to our actions. In other words, this is not an argument about some form of hard-line moral Realism which is somehow subconsciously related to and expressed by our automatic processes,
with the best way to hone it being a form of skilled practice. Instead, this is an argument that should lead us to think that the role of our conscious thoughts within action is complicated, and, that whatever their role turns out to be, it is at all times mediated by automatic processes when it come to actual actions. As I have made clear, general agentive goals can be expressed and deliberated about in detail. However, while these are one likely source of input, the actions themselves cannot be mediated other than through the automatic processes engaged at the time.

While there may be value to acts of deliberation, reasons for action are determined in the moment by the demands of a situation. From this we can wonder what an account of moral agency might look like. Contrary to the target Top Down Harmony (TDH) view attributed to Gendler in Chapter 4, I have argued two separate but interrelated points. First, following Brownstein and Madva (B&M), I have argued that there is a large category of actions, the nature of which is such that they could per force not be subject to an intention. From this we can conclude that there are automatic actions, and that the assessability of those actions happens in virtue of automatic processes. Second, following Arpaly and Schroder (A&S), I have argued that not only are there actions which could not possibly be subject to conscious planning, but that in the end acting for reasons does not happen in virtue of acts of deliberation anyway. Instead, acting for reasons happens in virtue of non-deliberative processes, that is, automatic processes. This is central for a palatable alternate account of moral agency. While B&M hold that automatic actions cannot be anything but norms responsive given that they are carried out in virtue of features too minute or quick to be available for rational reflection, A&S argue, and I have agreed, that we can think of
automatic actions as acting for reasons, in virtue of the content of the desires to which they respond.

We might wonder why it should matter that actions we would count as moral should be actions that happen for reasons, rather than behaviours that are merely norms responsive. First, deliberation is central to accounts of normative ethics because it is the primary mechanism through which reasons for action are identified. That we have an argument against the necessity of deliberation in order to act for reasons does not obviate the role of acting for reasons in ethical theory. What is really central for moral agency is acting for reasons, and that those reasons be good reasons. If we can preserve most of our normative ethical theorizing and only need to revise what exactly acting for reasons looks like on a human agent, then all the better.

The second thing to consider is whether B&M are right in not counting their own account of automatic action as acting for reasons. As we have seen, our reasons for action need not be apparent or available for conscious deliberation. Further, we have seen that it is due to an indefensible dichotomizing of conscious/unconscious as also rational/irrational, that we might think that processes which are not subject to conscious reflection are somehow non-rational. From Fridland, we have seen that automatic processes are intelligent all the way down, and from A&S we have seen that acting for reasons need not happen in virtue of acts of deliberation. While B&M give us reasons to think that features of automatic behaviour give us stand alone reasons to think that those actions are morally assessable, they fall short of a full take down of the TDH view of Gendler, falling
for the same sort of dichotomizing which underpins notions of moral actions as only morally assessable iff they cohere with our consciously avowed beliefs.

As a matter of future directions that might be interesting to explore, Fridland’s account of automaticity in combination with A&S’s view, can provide grounds to show that actions are happening in virtue of reasons irrespective of their availability to deliberation. A&S provide a desire based account of how this is the case. It is in virtue of the suite of intrinsic desires developed by a person that an action occurs. On A&S’s view, it is the conceptual content of intrinsic desires that does the work in providing reasons for action. If automaticity is cognitively penetrable, then we have the beginnings of an argument whereby the cognitive, conceptual content of intrinsic desires are responsible for action all the way down to motor control. One way to elucidate how this might be is through a focus on attention, and some of how A&S characterized the effects of intrinsic desires, and of good will, on attention In view of this, I take it as promising to think about A&S’s account of acting for reasons along with Fridland’s account of intelligent automatic action based on a picture of skilled behaviour.

As we saw in Chapter 2, work on attention is one source of evidence for the argument that automatic motor control is cognitively penetrable. As we saw with A&S, as intrinsic desires develop, attention is directed to the right cues, once an agent has developed good will. Thinking about how intrinsic desires develop attention, and how attention is tied to the fine grained motor control of, for example, saccades, can further help us characterize virtue as a sort of skill, and moral action as a sort of moral expertise. As I mentioned in the previous chapter, what is even more interesting about attention is that, as we saw in the Kentridge et al. experiments, visual attention can proceed non-consciously, which might,
on the face of it, be a surprising datum. If this sort of empirical data is correct, and if we agree with the characterization of agency I have provided thus far, then it is easy to see just how far from the evidence our current folk theoretical commitments lie. And, we can adumbrate the contours of an account of moral psychology that is enormously interesting and rife with possibility for investigation.

However, it is worth while noting that this is still not enough to take us away from the target Top Down Harmony view attributed to Gendler. After all, we can think that for the sort of skilled behaviour I am countenancing, some form of past deliberation (to use the terminology proposed by A&S) is sufficient for providing the cognitive content at the root of the intelligence that will trickle down to the finer grained motor array on Fridland’s view. Returning to attention, we can see how this might work on the account of moral action provided by Iris Murdoch. On her view, the moral agent’s worth is indeed revealed through the quality of their attention; that is, whether the agent is attending justly and lucidly to a situation of moral import so that they can then act as they ought to. However, the morally praiseworthy work, for Murdoch, resides in the sort of effortful imaginative and deliberative work that an agent can undertake in order to, for instance, better understand the point of view of an interlocutor she initially disliked. As we can see, if this is the sort of view of virtue we decide to subscribe to, then we have done little to vindicate the role of automatic actions as assessable in their own right.

If we pause and think about virtue on Aristotle’s account, we seem to find a similar problem for my purposes. Initially, the Aristotelian agent is essentially engaged in finding

---

maxims, and it is through the finding of the correct maxims for action (obedient to the doctrine of the mean), that an agent is to guide their action. This is, of course, a process of deliberation. The agent also is then supposed to become a moral expert skilled with the finding of the right virtue for the right time. This is important. It is not that once we have found the mean between ‘brave’ and ‘foolhardy’, this mean will serve us for all actions. It is not that once we have found a guiding principle for action, that this will translate into knowing how to act. We can think similarly about the role of normative ethics versus accounts of Moral Psychology. More importantly, this gives us room to push back against this maxim finding agent, or against the effortful, attention setting agent that Murdoch has in mind.

It is essentially a problem between translating the generality of a principle into the particularity of a behavioural space where one ought to act morally in the right way. Earlier I pointed out that B&M had come short of their own goal by falling victim to the same dichotomizing as Gendler when it came to assessing the rationality of an action. Recall, on their view, that they consider the sorts of automatic actions they have in mind too minute and specific to be the sorts of features which can be picked out in order to engage in a process of practical reflection, that is to say, deliberation. While I disagree that this means that it follows that those actions are thereby not actions done for reasons, I do think they have gotten the characterization of actual virtue exactly right. Moreover, if we recall A&S’s own account of virtue, this is much of what the moment by moment learning engaged by the reward system of the brain, which houses our intrinsic desires, turns out to look like. Returning to attention, on their account, the attentive agent of good will, will be attuned to all the right cues which enable them to act in a virtuous way. In this way, the particularities
of virtue as right action at a given time can be cashed out. Thinking of how B&M characterized the right attunement to a situation, I have already pointed out is very similar to how this looks for A&S. However, the emphasis that B&M place on micro-motor control provides an illustration of just how the sort of cognitive penetrability through an account of skill might proceed in virtue as a kind of expertise. If we account for virtue as skilled behaviour, where the finer grained control is responsive to the content of the intrinsic desires, then we have an account of automatic behaviour that is assessable on its own right precisely because it is beyond the reach of deliberation (for the reasons given by B&M), but which nevertheless counts as acting for reasons in a way that lets us preserve most of our current moral theorizing. Not only is this empirically attractive, but it is also intuitively appealing. The sort of kind demeanour I need to display towards a customer at my retail service job is not the same as kindness and concern, and the attendant emotional resources, I should call up when dealing with a grieving friend. Action, and thereby virtue, is particular. This account begins to make some sense of that, and of how to relate that to the sort of moral principles that normative ethical theorizing can elucidate for us.
Bibliography


Dreyfus, Hubert L., and Stuart E. Dreyfus. “What is Moral Maturity? Towards a Phenomenology of Ethical Expertise” In *Skillful Coping: Essays on the*


Gollwitzer, Peter M. “Implementation Intentions: Strong Effects of Simple Plans.”


Schurger, Aaron, Myrto Mylopoulos, and David Rosenthal. “Neural Antecedents of


