Considering the Options:
The Purpose and Authority of Practical Deliberation

by

Alexa A. Forrester

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2009

Doctoral Committee:

Professor Elizabeth S. Anderson, Chair
Associate Professor Sarah Buss
Associate Professor Gregg D. Crane
Professor Stephen L. Darwall, Yale University

To Dwight and Nita,
the dreamer and the pragmatist

Acknowledgements

I would like to thank the members of my committee: Liz Anderson, for the example you set, for all the time and energy you have invested in this project, and for the important resources you have provided intellectually and personally, including your apparently never-ending reserve of insight and calm. Sarah Buss, for the fact that you, joining this endeavor so late in the game, have somehow managed to provide a disproportional amount of priceless help. Stephen Darwall, for sticking with me from afar, and for being right about so many things. Scott Shapiro, for your general guidance, and for an incredibly helpful phone call at a pivotal moment in the development of my ideas. Gregg Crane, for deftly walking the line between philosophy and literature. I would also like to thank a number of other philosophers who have played a role in shaping my thoughts on this project: David Velleman, Rich Thomason, Eric Lormand, Daniel Stoljar, Peter Railton, Victor Caston, and Allan Gibbard. And, of course, I am grateful for the general support of others along the way, especially Louis Loeb, Ken Walton, and Herman Cappelen.

I would like to thank all of my fellow graduate students, and in particular David Dick, Tim Sundell, Josh Brown, Ivan Mayerhofer, and Eleni Manis. You know why I love you and what I owe you. I would also like to thank the many encouraging friends who have rooted for me along the way: Chris Cromwell, Heather Stickeler (Ikemire!), Liz Warburton, Derek Mong, Annie Fisher, Geoff and Kate and Iris Maturen, Vicki

Timpe, Erin Donnelly, Anne Schafer, Andrea Landis, and Alice Baker—I carry you all around with me every day. I would also like to thank the many amazing poets and writers who have, over the years, blessed me with beautifully-crafted breaks from philosophical texts.

I would like to thank my colleagues at the Association of American Colleges and Universities, who, during a moment of grad-school fatigue, reminded me why I love the academy. Michèle Leaman, I am especially grateful for your support as I literally grew into motherhood, and for your on-going friendship.

And then, of course, I would like to thank my family. Mom and Dad, for planting and nurturing the seed. To Jon, Bryan, Ginny and Cliff, Lani and Tommy, Sarah, Sandra, Gigi, Garth and Kendra, Andrea and Alison, Ann and Dave, and Charles, and to my extended family, Beth and Monte, Jean and Norm, Cindy and Mike, Frank and Laura, Barbara and Tom, and Melody, and all your above-average children and grandchildren, for ensuring that I always feel loved. Thanks to Teak and Cypress, my cats, for putting up with the multiple cross-country moves that were required to make this possible. To Cameron, thank you for your joyful disposition and patience. And, finally, to Chris, thank you—generally, specifically, and every other way—for every day so far.

Contents

Abstract

This dissertation is about practical deliberation: why we do it and what is at stake when we do. Contemporary ethicists tend to take one of two possible positions regarding the purpose of practical deliberation. Either deliberation is purely instrumental, serving heterogeneous ends fixed by motivational dispositions over which an agent has no rational control, or deliberation serves one particular meta-end—such as the systematic justification of desires or the inclination toward self-understanding—the possession of which is a precondition on rational agency. I argue that neither of these options yields an adequate understanding of motivational psychology, and offer instead a third option that takes deliberation to be a process of open-ended problem-solving aimed at overcoming actual instances of motivational uncertainty. I call this position the prospectivist account of deliberation.

Chapter One argues that prospectivism is a distinct theoretical option, reducible to neither instrumentalism nor a meta-end account of deliberation. Chapter Two defends the position that prospective deliberation meets the two constraints on practical reason: it is both practical and rational. Chapter Three argues that normative theory should not proceed without an awareness of the reflective conditions that give rise to the question of what to do. I call this the guidance-first approach to normativity. I answer Nomy Arpaly's explicit challenge to the guidance-first approach, and I use prospectivism to modify and correct existing accounts of the conditions that give rise to deliberation,

specifically, those offered by Thomas Nagel and Christine Korsgaard. Chapter Four defends an evidential theory of normative justification: a given value's capacity to guide unproblematic activity serves as evidence for, i.e. justification of, that value. Chapter Five argues that prospectivist deliberation suffices to explain the data that is usually cited in favor of free will, in part by providing an account of how prospectivist deliberation can be used to justify claims of moral responsibility. Since prospectivist deliberation is also compatible with determinism, this effectively deflates the 'problem of free will'.

Chapter One

Why Deliberate?

1.1 The Desiderata

In order to support theories of agency and practical reason, our best account of motivational psychology must allow us to do the following.

(a) Ground the distinction between the total set of an agent's physical behavior and the subclass of her behavior that is motivated behavior (i.e., goal-oriented behavior).

(b) Explain the trigger for, and aim of, deliberation. This includes accounting for both means-end deliberation and deliberation of ends (reflection during which we consider potential ends themselves and attempt to determine whether to adopt those ends).

(c) Ground the distinction between motivated behavior and the even smaller subclass of deliberate behavior.

These desiderata are interrelated. We should expect that how we account for any one will influence how we account for the others.

1.2 Standard Belief-Desire Psychology

The position that belief and desire are distinct mental states, and that the former is governed by reason, while the latter is not, is often attributed to Hume and called 'Humean.' I wish to sidestep interpretive issues here, so I will focus on the position itself, which I will call "standard belief-desire psychology." The basic idea is that motivated behavior is caused (and distinguished) by two isolated components—a

volitional component and a cognitive component. The former fixes the ends at which an agent aims, and the latter fixes the means to those ends.

Belief-desire psychology can easily ground the distinction between mere behavior and motivated behavior. Any behavior guided by appropriately related belief-desire pairs counts as motivated behavior. Any behavior not guided by beliefs and desires does not. But the belief-desire model encounters difficulties when explaining the trigger for, and aim of deliberation, and grounding the distinction between all motivated behavior and a smaller subclass of consciously-guided deliberate behavior.

One initial problem is that, if passion alone is responsible for fixing ends, then it seems we are faced with a dilemma. First, we might believe passion is determinate—that whenever there is a case of conflict between desires, the intrinsic properties of the desires is what settles the issue.[1] The problem with this is that it would make the familiar phenomenon of practical deliberation about goals inexplicable. If an agent's desires are on their own equipped to resolve any potential conflict with other desires, then an agent would never have cause to deliberate about whether she should do one thing or another. And if she did so deliberate, such deliberation would be causally inefficacious. Her passions would dictate what she was going to do, unresponsive to her judgment of whether she should do it. This hypothesis contradicts our common experience of instances in which we try to decide between two competing goals by reflective deliberation, instances in which our decisions seem, at least in most cases, to make the difference in what we do. Of course, our experience may be deceptive. Perhaps it simply

[1] One popular candidate property to play this role is brute hydraulic strength of a desire, but some other property could serve the role instead. The important point is that, whatever property serves the role, it must be intrinsic to the competing desires, and capable of settling the competition on its own, i.e., without arbitration or moderation.

seems like we engage in causally efficacious deliberation of ends, but in fact we never do. This is a conclusion, I believe, we should accept only if and when we have ruled out all competing theories of motivation that would vindicate our deliberative experience.

The second prong of the dilemma is to admit that while passion is the only thing that can fix one's ends, sometimes it fails to do so. And in these cases, we become unsure of what we will do. But this cannot be right either. For, although it preserves our intuition that we can sometimes be uncertain about what to aim for, it gives us no power to overcome this uncertainty. In other words, it does not help explain the trigger or aim of deliberation, only the cause of uncertainty. Furthermore, if this were an accurate account of motivation, we would expect to encounter much more often than we actually do cases in which agents have become motivationally stuck. All around, our fellow citizens would be simply standing still because their passions were in conflict, were not themselves capable of resolving the conflict, and nothing else was.

Of these two options, the first is more tenable. Therefore, a regular companion of belief-desire psychology is hydraulicism: the claim that the stronger desire (in the hydraulic, rather than phenomenological sense of strength) always wins, and that in almost all cases of conflict one desire is stronger than the other.[2]

Hydraulic belief-desire psychology can straightforwardly explain the distinction between motivated and mere behavior. To have a goal is to have an end that is backed by a desire stronger than any competing desires and enabled by appropriate means-end beliefs. Motivated behavior is all behavior guided by belief-desire pairs; mere behavior

[2] Again, hydraulicism is only one option for how passions might be causally determinate. I will not explore other possibilities here, except to say that any theory that proposes that desires are always capable of working things out on their own will be subject to the same criticisms I level against hydraulicism.

is any behavior not guided by belief-desire pairs. But what resources, if any, does hydraulic belief-desire psychology have to account for the phenomenon of deliberation, particularly deliberation of ends, and the difference between the broad class of motivated behavior and deliberate action?

One approach to explaining desiderata (b) and (c) within the standard belief-desire framework is to posit a meta-end that 1) triggers deliberation, 2) is not itself in question during deliberation, 3) provides standards whereby other ends are organized, and 4) enforces the conclusions of deliberation. The literature offers many candidate meta-ends, for example: the aim of doing what is best all things considered (Davidson[3]); the aim of maximizing utility (rational choice theory); the aim of acquiring self-understanding (Velleman[4]); the aim of doing what is justified (Smith[5]); and the need to act only in accordance with reasons for acting (Korsgaard, on one interpretation[6]). Granted, these meta-ends are not always offered specifically in response to the above desiderata, nor as meta-ends as such. However, I believe it is the dialectical responsibility of those who posit these deliberative aims to explain how else they can be interpreted, if not as meta-ends.[7]

During deliberation, a meta-end has authority not because of its strength, but because of its status. As the end whereby other ends are organized, i.e., the end that

[3] Davidson, Donald. 1980/2001. See, in particular, "How is Weakness of the Will Possible?" (21-42)
[4] Velleman, J. David. 2000.
[5] Smith, Michael. 1994.
[6] Korsgaard, Christine. 1996.
[7] Another option for explaining deliberation of ends is to posit two distinct sources of motivated behavior, one that is entirely accounted for by hydraulic belief-desire psychology, and another, which operates entirely independent from the influence of desire and which generates intentions and choices. This approach is taken by R. Jay Wallace (2006). I believe this position is unstable, fluctuating between either positing a hydraulically backed meta-end or endorsing an untenable metaphysics.

deliberation serves, it is not itself in question. Of course, in order for a meta-end to trigger deliberation in the first place, and to enforce deliberation's conclusions, it must be backed by a volitional mechanism that is stronger than any of the various heterogeneous desires that would otherwise move the agent. By attributing to all deliberators a meta-end backed by a such a mechanism, we can account for the existence of agents who deliberate regularly about what their goals should be, while still maintaining the spirit of belief-desire psychology.

Turning to desideratum (c), we must ask how a meta-end might help explain the difference between motivated behavior and deliberate behavior. To illustrate the difference, consider the following example from Velleman.

> …a child accidentally brushes a glass off of the table, and your hand shoots out to catch it. Everything happens so fast that you see your hand catching the glass before you fully realize that the glass is falling…[This] is an instance of behavior directed at a goal, but it isn't a full-blooded exercise of your agency…the reflexive extension of your hand is aimed at something—namely, preventing the glass from smashing on the floor. Despite being goal directed, however, this behavior still lacks some element that's necessary to full-blooded action. So what makes for action is not simply being goal directed.[8]

How can belief-desire psychology make up the difference? Again, meta-end accounts can be deployed here. The requisite distinction between automatic goal-oriented behavior and deliberate goal-oriented behavior can be grounded in an agency-bestowing meta-end. For example, Velleman takes any action guided by the aim of self-understanding to be full-blooded action.[9]

In what follows, I hope to show that the three desiderata can be better explained without positing a meta-end. The explanation will be better in the sense that it is more

[8] Velleman, J. David. 2000. (189-190)
[9] Velleman, 2000. (See in particular 193-199)

parsimonious and also because it preserves some room for a genuine sense in which we, rather than passions or inclinations that we happen to have, control our fates.

1.3 A Prospectivist Account of Motivated behavior

1.3.1 Deweyan starting points

John Dewey offers an understanding of the cause and purpose of practical deliberation that differs significantly from the meta-end view. He begins with the observation that humans are naturally active beings. Second, he observes that human activity[10] is guided primarily by habit and impulse. In other words, most of what we do is done not because we decide to do it, but rather because we have developed a habit of doing it, or are moved by some subconscious impulse to do it. This does not mean that there is no explanation for such activity, just that the explanation does not involve an appeal to conscious desires and aversions, nor to the process of deliberation. Which brings us to the third salient feature of Dewey's theory: It is only when a conflict among habits, impulses, and/or the environment interrupts our natural activity that deliberation begins. That is, deliberation is a means to restoring activity that, until a particular point of conflict, was proceeding smoothly. The fourth important point follows naturally: the constitutive aim of deliberation is the overcoming of conflict and the subsequent restoration of natural activity.

Dewey believes that humans engage in "original, unlearned activity."[11] His name for the forces that spur this activity is impulse. Infants emerge from the womb crying, clenching their fists, kicking their legs, etc. It seems unnecessary to posit any directive

[10] This term is meant to be neutral, encompassing both mere behavior and action.
[11] Dewey, John. 1922/2002. (92)

mental states to these infants. We do not believe their crying is the result of a decision to cry, or even a consciously-entertained desire to do so.

Dewey emphasizes that, to a greater extent than most other creatures, humans are born dependent on elder members of the species. This fact interests him because, as a consequence, an agent's survival and prosperity requires her ability to immediately begin harmonizing her impulses with both the physical and the complex social environment into which she is born. Dewey writes, "Even if by some miracle original activity could continue without assistance from the organized skill and art of adults, it would not amount to anything. It would be mere sound and fury."[12] Thus, harmonizing our impulses with our social and physical environments is not only necessary for survival; it is the source of the meaning of our activity.

I find it provocative to consider the notion that all meaning, including linguistic meaning, might arise from the activity of harmonizing our native impulses with environmental stimuli by way of habit formation. But for now I will stick with a more modest and less controversial interpretation of Dewey: after so much hapless flailing, an infant begins to notice that every time she places her legs at a certain angle, they connect with the floor and her whole body moves.[13] The act of placing her legs at a certain angle while kicking, in virtue of its association with her subsequent movement, thus takes on significance, or comes to mean something to her. In this sense, agents naturally expect like effects from like causes. In the highly socialized, linguistic environments in which most human children are raised, the muttering of a certain sound can come to mean the appearance of a comforting blanket; a certain expression on the face of a parent can come

[12] Dewey, John. 1922/2002. (90)
[13] I do not mean to suggest by this that the infant is yet aware of her legs as her legs. Rather, simply that she is aware of predictable patterns of felt activity and felt response.

to mean impending unpleasantries; the tone of a certain buzzer can come to mean that someone is at the door. Meanings of this sort are inextricably tied to an agent's learned expectations and her preference for certain courses of events over others.

The process of harmonizing our natural impulses for movement and expression with our environment leads to the formation of habits—habitual behavior that allows for the regular release of impulse in a manner that responds to both social and natural environmental pressures. For example, once an agent learns that sucking on a pacifier is an efficient way to release a particular impulse (as opposed to sucking her thumb, which invites troubling environmental factors like disapproval from her elders), she develops the habit of sucking on her pacifier. In this way, we develop habits through which our impulses are channeled.

For Dewey, there is nothing more to an agent's will than the total collection of the habits that move her to act. He writes, "All habits are demands for certain kinds of activity; and they constitute the self. In any intelligible sense of the word will they are will."[14] It is important to acknowledge that Dewey's definition of will diverges significantly from the definition that prevails in most philosophical discussions, the definition upon which will is embodied in deliberate action, in action on the conclusion of a piece of conscious practical reasoning. It may seem puzzling that Dewey chooses to include under the canopy of "will" operations that happen unconsciously or behind our backs. But I hope it will become clear as I proceed that for Dewey, almost everything we do is determined in one way or another by habit, including reflective, intelligent conduct. Taking this into consideration, we can see that Dewey's equivalence of an agent's will

[14] Dewey, John. 1922/2002. (25) Emphasis his.

with her habits and impulses constitutes a definition that is broader but not contradictory to the one normally invoked by philosophers.

Dewey believes that very little of our activity is the result of pure, unshaped-by-habit impulses. In the following passage, his discussion of the persistence of certain societal customs helps clarify his sense of the relationship between impulse and habit.

> War and the existing economic regime…are crucial cases of the relation existing between original impulse and acquired habit. They are so fraught with evil consequences that any one who is disposed can heap up criticisms without end. Nevertheless they persist. This persistence constitutes the case for the conservative who argues that such institutions are rooted in an unalterable human nature. A truer psychology locates the difficulty elsewhere. It shows that the trouble lies in the inertness of established habit. No matter how accidental and irrational the circumstances of its origin, no matter how different the conditions which now exist to those under which the habit was formed, the latter persists until the environment obstinately rejects it. Habits once formed perpetuate themselves, by acting unremittingly upon the native stock of activities [i.e. impulses]. They stimulate, inhibit, intensify, weaken, select, concentrate and organize the latter into their own likeness. They create out of a formless void of impulses a world made in their own image. Man is a creature of habit, not of reason nor yet of instinct.[15]

For reasons illustrated in this passage, Dewey criticizes the projects of psychologists who attempt to classify our "original impulses" into rigid categories such as the impulse for sex or the impulse for war. Impulses, he believes, are non-specific, and they secure their release in accordance with whatever means are available to them. For this reason, individual habit and societal custom persist because they provide a ready outlet for a particular impulse, but not because the impulse itself demands that particular outlet. The concluding sentence of the above passage supports a claim I made earlier. Dewey believes that a large portion of human activity results from the impulses humans have and the habits through which those impulses are regularly released. That is, a large portion of

[15] Dewey, John. 1922/2002. (125)

human activity is what contemporary philosophers would call mere behavior, as opposed to deliberate behavior or action.

For Dewey, the need for deliberation arises only when an agent's natural activity is arrested—when habits and impulses are incapable of moving the agent either because they have come into conflict with each other or with the environment. When this happens, the agent becomes temporarily paralyzed and deliberation serves as the means for overcoming this paralysis. For this reason, it is proper to say that resolving the impasse among habits, impulses, and the environment is the constitutive aim of deliberation. During deliberation an agent is quite literally being propelled in two or more directions at once, or in one environmentally unavailable direction. The end of deliberation is the cessation of this tension, the return to a state in which the agent is being propelled in one and only one (available) direction.

The question remaining is how an agent deliberates. What process does an agent use to overcome the impasse she is confronting? According to Dewey, to overcome this temporary paralysis, she must actively seek a resolution by imagining a variety of possible actions and their respective consequences in search of one imagined option that will, in virtue of its representational content, restore overt action. Dewey writes: "…deliberation is a dramatic rehearsal (in imagination) of various competing possible lines of action,"[16] and, "All deliberation is a search for a way to act…[its] office is to facilitate stimulation,"[17] and finally, "[Deliberation] consists in selecting some foreseen consequence to serve as stimulus to present action."[18]

[16] Dewey, John. 1922/2002. (190)
[17] Dewey, John. 1922/2002. (193) Emphasis his.
[18] Dewey, John. 1922/2002. (261)

The process of deliberation terminates in choice, the event of "hitting in imagination upon an object which furnishes an adequate stimulus to the recovery of overt action."[19] In other words, choice occurs when the agent imagines an available course of action that satisfies the demands of her will.[20] This particular imagined course of action resolves the temporary paralysis of conflict, and allows overt action to proceed. On this account, choice is envisioned as a type of discovery, the terminus of a specific search—the search for a unified path forward. Dewey writes, "There is seen to be but one issue involved in all reflection upon conduct: The rectifying of present troubles, the harmonizing of present incompatibilities by projecting a course of action which gathers into itself the meaning of them all."[21]

Dewey's understanding of motivation and deliberation have been mostly overlooked or dismissed by contemporary ethicists, I believe, to the detriment of the field. One of my primary aims in the coming pages is to resuscitate key points from his work in the context of a new theory of motivation and deliberation, and to demonstrate the power and intuitive appeal of such a theory. In the process, I will diverge considerably from Dewey's terminology and some of his theoretical commitments, but the motivational psychology I develop below is much indebted to those aspects of his theory highlighted above.

1.3.2 A new way to think about motivated behavior

To develop these ideas in greater detail, I propose we begin by setting aside the idea that causally efficacious belief-desire pairs guide all motivated behavior and

[19] Dewey, John. 1922/2002. (192)
[20] Again, for Dewey, will is equivalent to the total collection of an agent's habits and impulses.
[21] Dewey, John. 1922/2002. (210)

replacing it with a single mental state—expectation—that causes and guides motivated behavior. I do not deny that both beliefs and desires play a role in fixing the content of expectations, but illuminating how this happens is key to understanding our motivational selves.

For the purposes of this thesis, I will define expectation as the future-oriented epistemic state (whatever it is) that enables surprise. For example, imagine an agent walking down a hallway in the middle of the night and tripping on a box she had left there the previous evening. As this agent walks, she does not represent the presence of this box to herself, and instead is guided by the expectation that she will, as she normally does, walk safely down the hallway. Thus, she is surprised to find resistance against her foot as it collides with the box, surprised to find herself suddenly thrown off balance, etc. Now, to say that, prior to the collision, she believed there was no box in the hallway, or that she believed the hallway to be entirely clear, would be strained, since she had not reflected on either of these propositions, and if she had, she probably would have changed her expectations. Nevertheless, she was expecting the hallway to be clear, and hence expecting herself to walk safely ahead, and, I propose, this expectation was guiding her behavior.

Why think that expectations cause behavior? Consider first our simplest expectations, those that, like perception, are generated automatically in response to the environment. The psychologist Daniel Gilbert calls this phenomenon, in which an agent's mind regularly and automatically generates expectations of what's ahead, nexting.[22] Gilbert cites an experiment in which monkeys watch a ball drop into a chute. The monkeys automatically lower their eyes to the bottom of the chute in anticipation of

[22] Gilbert, Daniel. 2006. (7-8)

the ball's reemergence. Gilbert writes: "When some experimental trickery causes the ball to emerge from a different chute than the one in which it was deposited, the monkeys display surprise, presumably because their brains were nexting."[23]

But if nexting can cause small behavioral impacts like the redirecting of one's gaze, as in the case of the monkeys, then there is no principled reason that it could not cause impacts in larger muscle groups as well. In fact, this seems like a good explanation for what happens when, say, a cat moves its whole body in anticipation of the future location of a particular mouse. Or consider this: Imagine you are waiting at the bottom of a slide to catch a toddler. Suppose one section of the slide is covered, so that, momentarily, the child disappears from view. Without any conscious effort, your brain will, before she reappears, generate the expectation that she will reappear, including a rough expectation of the velocity and direction she will be traveling. This expectation will guide not only where you look, but also how you position your body.

But notice also how this last scenario is more sophisticated than the monkey case. The representational content of your automatic prediction of the child's trajectory is not working alone to guide your behavior. There is another structure of representational content, namely the image of yourself catching the toddler that influences your behavior. We can demonstrate this counterfactually. Replace the toddler in the example with a snarling terrier. Although your expectations of the terrier's trajectory might be roughly the same as in the case of the toddler, the effect of that expectation on the position of your body will be quite different.

So there is yet another representational structure influencing your behavior, an activated (or practically engaged) representation of catching the child in the near future.

[23] Gilbert, Daniel. 2006. (8)

Which is to say, the temporally-indexed contents of your expectations are shaped not only by current perception and background beliefs, but also by the content of some activated representation of the future, which I will call a goal.[24]

Even very sophisticated goals can be automatically and unconsciously generated. In his article, "The Unbearable Automaticity of Being," psychologist John Bargh summarizes a variety of studies which demonstrate that nearly all behavior—from the simplest movement of one's eyes, to much more sophisticated, goal-oriented behavior like aiming to keep a child safe—can be governed by goals activated without an agent's awareness, and without a conscious act of will.[25] In one example, subjects could be made to adopt unknowingly the goal of excelling on a test simply by being exposed to the word "success" in the course of an earlier unrelated task. Bargh concludes that, even if a small percentage of goals are activated by what he calls "acts of will,"[26] most human behavior is governed by automatically activated goals. For now, the important claim is that these activated goals, however they become activated, shape our expectations of what is ahead, including our expectations of what behavior we will engage in and what will result from that behavior.

---

[24] I am using this term in the way it has been employed in psychological literature on this topic. As I will soon discuss, it is broadly defined to include both automatic, unconsciously activated goals and more deliberately adopted goals.

[25] Bargh, J. A., and Chartrand, T. L. 1999.

[26] Bargh characterizes behaviors governed by acts of will in a variety of ways: as processes that we choose and intend to engage in; processes of which we are aware; processes that require effort and attention; and processes that we can control. From a philosophical perspective, if this list is meant to distinguish "automatic" from "willful" behavior, it is far too messy. Each of these descriptions specifies a slightly different set of conditions, leaving unresolved the question of which conditions are essential to, and which are merely symptoms or indicators of, acts of will.

1.3.3 Prospects, prospections, and the prospective set

Suppose we think of this in the following terms. Call any representation p whose content may become expected a prospect. Prospects come embedded in larger representational structures, strings of temporally-indexed expectations that represent the activity leading up to, resulting from, and including a particular goal, ϕ. Call each of these larger structures a prospection-aimed-at-ϕ, P_{ϕ}. This concept is meant to capture consciously adopted long-term plans, as well as more short-term and sometimes automatic expectations.

The contents of prospections are extremely rich, dynamically evolving, and they can draw on any concepts in an agent's conceptual arsenal. An agent can, for example, expect to buy a motor home, to participate in democracy, to overcome her fear of flying, or to eat a grape. Prospective content is in large part determined automatically in response to an agent's perceived environment and to the specific contents of her goals as she is representing them. Prospective content may also, if necessary, be determined via deliberation, but I will return to this point later.

At the foundation of prospective moral psychology is the claim that all motivated behavior is caused by prospections. Let me say a bit more about this causal claim. First, it is only an agent's expectations of what will happen in the immediate future, rather than the entire contents of her prospection, that actually trigger the motor responses that now cause overt behavior. And yet the agent's representations of the temporally immediate future are influenced by representations of what will happen in more temporally distant moments. So the activated goals in prospection play a crucial, if indirect, causal role in guiding current behavior.

Second, prospections come in varying degrees of specificity. The expectation that I will make myself a cup of coffee does not include a representation of the specific movements I will make with the muscles in each of my fingers. This is important because it turns out that the behavior caused by a given prospect in prospection is a lot more intricate than is explicitly represented. Skilled behavior, I maintain, is just this: representations at a general level causing very intricate behavior that satisfies the general representation.

Third, we are not simple creatures. Often, our behavior is guided by many prospections at once. We can thus speak of an agent's entire set of prospections, or her prospective set.[27] And, to return to desideratum (a), we can define motivated behavior as the class of all behavior guided by the content of an agent's prospective set.

1.3.4 The inaccuracy of hydraulicism

By positing that the contents of an agent's prospective set, including the goals therein, often evolve dynamically and without reflection, prospectivism accommodates Bargh's empirical discoveries, and our personal experience, both which suggest that much of our goal-oriented behavior is automatic, in the sense that it is triggered without conscious reflection or awareness. This brings us to the task of explaining (b)—what causes us to deliberate and what is the aim of our deliberations?

Prospectivism offers the following account: allowing that an agent's prospective set can evolve automatically and without disruption, there are frequent cases in which an

[27] Determining the content of an agent's prospective set at any moment is at least as difficult as the notoriously difficult task of determining what an agent perceives at any given moment. Nevertheless, we can easily evaluate certain claims about content. I.e., I am perceiving a computer screen. I am expecting my keystrokes to cause words to appear on the computer screen. Etc.

agent's prospective set becomes unsettled, i.e., when some change in the agent or her environment automatically activates new prospective content that cannot be seamlessly incorporated into her current prospective set. When prospection is thus interrupted, an agent literally does not know what to expect from herself, and because of this, 1) she is not prompted by her expectations to do anything[28], and 2) her future behavior seems indeterminate to her.

Importantly, this can and does happen on its own. Interruption of prospection is all that is needed to trigger deliberation, and we find ourselves needing to deliberate precisely because automaticity fails us, because our heterogeneous motivations and beliefs as they are at a given moment do not add up to settled prospection. In such cases of disruption, in order to restore motivated behavior, we need to decide what to do. Practical deliberation is the process employed to overcome such impasses, and choice is the conclusion of that process. Thus, on the prospectivist account, agents deliberate regularly not because a particular meta-end requires it, but rather because agents are equipped with an automatic deliberative response in the face of prospective uncertainty.

But this raises a number of further questions. How does prospection become unsettled in the first place? Is the aim of settling one's prospective set itself a meta-end? What is required to resettle prospective content? What deliberative processes are involved? I turn to these questions next.

[28] At least not with respect to the particular prospections involved in the interruption. Prospection can be partially interrupted.

1.3.5 Three causes of unsettled prospection

All cases of unsettled prospections, I maintain, are caused by one or more of three possible problems with prospective content: the perceived presence of a gap, the perceived presence of contradictory content, or the motivational insufficiency of content.

Gaps: A gap in prospection occurs when an agent is unable to generate content that she can expect to get her from her present position to some activated goal. For example, suppose an agent is aiming to reach a certain campsite by nightfall. To a certain degree of specificity, she expects a future in which she continues hiking on the trail and arrives at the campsite before sunset, and this representation is guiding her conduct. Now suppose she comes upon a raging creek that, because of flash flooding, has washed out the footbridge. This causes a gap in her prospection: she can no longer expect herself to continue on the trail, and there is no other course of activity she can imagine herself engaging in that she can expect will deposit her at the campsite by nightfall. Thus, the timeslots in prospection between the representation of the present and the representation of her pre-dusk arrival at the campsite are rendered devoid of expectable content. I will call gap-free prospections <u>means-end contiguous</u>.

Gaps can be caused and perpetuated by lack of means, lack of skill, or lack of imagination. It may be, in the case above, that there is no human that could ford the river in those conditions. In such a case, the agent's inability to generate a settled prospection stems from an accurate representation of her lack of means. Similarly, it might be that the hiker could ford the river with appropriate training, but that she rightly gauges she lacks such training. In this case, the

hiker's accurate understanding of her lack of river-fording skill causes the gap.[29] Or, finally, it might be that the hiker could ford the river by simply trekking upstream, where the river is preceded by two smaller, fordable creeks. In this case, the hiker has the required means and skills needed to reach the campsite by nightfall, but simply fails to realize this. Here, her failure to imagine something that is in fact possible perpetuates her impasse.

Importantly, there is room for uncertainty and experimentation in the face of a problem. She can, for instance, adopt an interim goal of stepping into the river to test its strength before determining whether she can expect to ford across. And there is also room for hopes and ideals—outcomes an agent aims for but doesn't actually expect—to influence prospection. Consider an individual rushing to catch a bus. She may not expect to make the bus (i.e., she will be disappointed but not surprised if she sees the bus pulling away as she arrives at the stop), but she does expect to get to the bus stop quickly. The prospection aimed at getting to the bus stop quickly will involve running, crossing streets without waiting for a walk signal, etc., until she reaches the bus stop. Her activated goal is shaped by her hope that the bus will be there when she arrives, but the object of her hope is not itself expected.

Contradictory content: Just as it is impossible to consciously and simultaneously endorse p and not-p, it is impossible to consciously and simultaneously expect p and not-p. Of course, we can sometimes fail to notice contradictions in our expectations. I can on Monday adopt a plan to meet with Dave on Wednesday night, and

[29] Lack of skill can be seen as a subset of lack of means, but one that is useful to distinguish when it comes to diagnosing our problems and planning long-term preventative measures.

on Tuesday agree to meet with Ivan on Wednesday night. In the intervening time, I may fail to think about the two contradictory events simultaneously.[30] More likely than not, however, sometime prior to the planned meetings I will realize that I am expecting two contradictory things, and this realized contradiction will unsettle the contents of my prospective set. When this happens, in order to resettle my prospective set, I must either surrender (i.e., deactivate) one of the goals, or find some way to reconcile the two.[31] Unfortunately, these processes can be very difficult, especially when we are strongly attached to certain outcomes.

Motivational insufficiency: Imagine you are rushing to leave your house when your housemate pulls a fresh sheet of piping hot cookies from the oven and invites you to have one. Your mouth starts watering. Your mind generates an image: head to the kitchen, reach out, grab a cookie, lift it to your mouth, etc. This whole chain pops into your head automatically. However, your knowledge of the cookies' temperature generates the expectation of pain at the end of such a sequence. And so, in spite of your possession of an epistemically viable prospection, your motivational system balks. You

[30] My failure to recognize the conflict can be due to the influence of what Rich Thomason has called "topic." Our representations are often circumscribed by topic to help us plan more effectively, but as Thomason notes, distortions due to topic are "one price we pay for such circumscription." (See, Thomason, Richmond. "The Context Sensitivity of Belief and Desire" in Reasoning About Actions and Plans. Georgeff, Michael P. and Lansky, Amy, ed. 1987)

[31] There is actually a third option: I can willingly ignore the contradiction, putting it out of my mind, especially if it occurs in some temporally-distant time slot. This will allow motivated behavior to continue in spite of the contradiction. However, as the moment of contradiction grows closer and closer, it will get harder and harder to do this, for if the contradiction occurs in the most immediate timeslots of prospection, it will disable current behavior: I literally won't know what to expect from myself, and I won't be able to do anything until I resolve the conflict. As Bratman has argued, humans are planners by nature. We tend to work out the kinks in prospective content before the time for execution is upon us. However, there may be individuals who chronically wait to decide. They foresee future incompatibilities and pernicious gaps but ignore them until they are literally forced to reconcile them.

consider waiting for them to cool, but your motivational system also balks at being late leaving the house.

This example highlights the fact that epistemic viability is not the only constraint on the expectability of content. Excitement and aversion are the fuel of all motivated behavior; they are the springs of action. If our expectations of the future did not physically excite or avert us, they would not cause motivated behavior at all. And this yields the third possible source of prospective interruption: motivational insufficiency.[32]

It is worth noting that empirical research suggests that neurological systems of reward and punishment are distinct.[33] My strong aversion to the representation of being burned is not matched by a strong excitement at the thought of not being burned. Nor must my strong excitement at the thought of eating a cookie be matched with a strong aversion to representations of not eating a cookie. We need not try to reduce aversion to desire, nor reduce interruption due to motivational impotence to interruption due to incompatibility of content. Faced with the likely prospect of getting hurt if I grab a cookie off the cookie sheet, we need not suppose that the cause of my hesitation is due to the fact that getting hurt conflicts with an active goal of not getting hurt. It is possible and advantageous to keep the conditions for epistemic viability separate from the conditions for motivational viability.

---

[32] This may seem trivial, but the condition is actually informative. First, it is only one of three possible sources of interruption, so a prospection that is extremely exciting can fail to cause behavior if it is lacks means-end consistency or compatibility. And prospections that are both means-end consistent and compatible can fail to guide behavior if they are not exciting enough (including the case in which they are positively averting). Second, being sufficiently exciting is a matter of whether the content represented in prospection generates a certain threshold-level of excitation, and a given prospection can cross this threshold in a number of different ways.

[33] Timothy Schroeder offers a helpful summary of the neuroscientific research on punishment and rewards in his book Three Faces of Desire. 2004. (49-57)

The motivational impacts of various representations are determined in part by nature, but also a great deal by nurture, particularly by the cultural environments in which we acquire our concepts. Consider, for example, how much the cultural context in which one acquires the concept 'abortion' affects the motivational responses one has to prospections involving that concept. Because these motivational impacts are physical impacts, they are, in principle, measurable. But I leave the task of discovering how to measure them to the empirical psychologists.

To summarize, in order to be incorporated into settled prospection all potential prospective content must simultaneously satisfy three constraints, two of which are epistemic, and one of which is motivational. In this limited sense, prospectivism mimics belief-desire psychology. But it replaces belief-desire psychology's bifurcation with the notion of a single mental state—a prospection—the entire represented content of which must be both epistemically and motivationally viable in order to guide activity. The novel claims of a prospective moral psychology are a) any automatically-activated content that, as represented, satisfies these three constraints, is incorporated seamlessly into settled prospection, influencing, but not interrupting, an agent's motivated behavior; and b) in the case that some automatically-activated prospective content fails to satisfy these three constraints, an agent's prospective set becomes unsettled, and practical deliberation is triggered.

1.3.6 The aim of deliberation

Prospectivism maintains that agents are compelled to deliberate about what to do because deliberation is how they actively bridge the gap between a state of suspended motivation and a settled prospective set. This offers a clear contrast with the meta-end

accounts discussed above. Consider the descriptions of deliberation found in the following passages:

From Christine Korsgaard:

> The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, as long as it reflects, it cannot commit itself or go forward…[O]ur reflective nature gives us a choice about what to do.[34]

From R. Jay Wallace:

> In deliberation, we try to determine what to do by reflecting on the question of what there is reason for us to do…The power of choice has its natural home in…the deliberative question of what we ought to do.[35]

From J. David Velleman:

> One possibility [for a motive that would bestow rational agency] would be to posit a higher-order desire, on the part of every agent, to be actuated by those of his lower-order motives which constitute the best reasons for him to act. This desire would move a person to survey and evaluate his motives as reasons for acting, and it would then add its motivational force to whichever combination of motives impressed him as rationally superior.[36]

From Michael Smith:

> Suppose Anne very much wants to dance a jig. Recognizing her desire to dance a jig, she asks herself whether this desire is worth having and acting upon. In other words, she deliberates…When we deliberate, and decide what we have a rational justification for doing, that very fact sometimes makes a difference to what we do.[37]

All of these writers depict agents who are, first, determinately motivated by belief and desire, but who then, in an act expressive of their rational nature, survey their desires (or the options suggested by their desires) and ask, "I want to do x, but should I?"

But deliberation almost never happens this way.[38] Not only are our desires <u>as such</u> rarely in the foreground of deliberation, but also, most of the time we deliberate, we

[34] Korsgaard, Christine. 1996. (92-93, 96)
[35] Wallace, R. Jay. 2006. (151, 153)
[36] Velleman, David. 2000. (14)
[37] Smith, Michael. 1994. (153, 132)
[38] Nomy Arpaly makes a similar point. Arpaly, Nomy. 2003. (24)

simply are not motivated, consciously or otherwise, by some latent love of or need for reasons. We are motivated, rather, by the very things that have caused conflict in our prospective set, and we struggle to move forward in the face of the conflict. A prospective moral psychology captures this fact. On a prospective foundation, the constitutive aim of deliberation is nothing more substantive than deciding what to do. This is the minimal aim all practical deliberators have in common.

1.3.7 Is 'deciding what to do' a meta-end?

Christine Korsgaard writes,

> Human beings are condemned to choice and action. Maybe you think you can avoid it, by resolutely standing still, refusing to act, refusing to move. But it's no use, for that will be something you have chosen to do, and then you will have acted after all…The necessity of choosing and acting is not causal, logical, or rational necessity. It is our plight: the simple inexorable fact of the human condition.[39]

I take Korsgaard to be making the following point here. Once you have donned the practical point of view, there is no way to proceed with motivated behavior without first making a choice (ignoring, for the moment, cases of shifts in circumstance). Prospectivism accommodates this point. To see how, consider the following example.

Suppose you must decide by tomorrow between two competing job offers, one in Boston and one in San Francisco. After spending a great deal of time weighing the pros and cons of each, you have yet to settle the contents of your prospective set. Under such circumstances you might be tempted to throw up your hands and say, "Forget it! I simply cannot decide!" The problem with this is that once you have donned the practical point of view, once the contents of your prospective set have become unsettled, the option of

[39] Korsgaard, Christine. 2009. (2) Her emphasis.

not engaging in practical deliberation is just that—an option. And thus it must itself be perceived as means-end contiguous and non-contradictory, and it must be motivationally sufficient. But notice, the fact that you are strongly excited by the prospect of accepting the Boston job works against the prospect of making no decision as much as it works against the prospect of accepting the job in San Francisco.

This is how a conflict in motives can itself motivate deliberation. The goals that cause the problem continue to make their demands until you, the deliberator, figure out some way to manage them. And, from the deliberative prospective any hypothetical prospective content, including the option of suspending deliberation without making a decision, or the option of deciding what to do by flipping a coin, is sensitive to the motivational and epistemic constraints that govern all prospective content.

This is why, as Korsgaard correctly observes, you cannot avoid choice and action. Because if you are ever in a situation in which you are seriously considering not making any choice, it is too late. The responsibility for your future behavior already thereby lies in your hands (or rather, in your deliberative processes, which I discuss more below) and, short of spontaneous suspension of your mental faculties or abduction, nothing is going to relieve you of that responsibility. But, importantly, prospectivism accommodates this necessity of choice without committing us to any particular ends, meta- or otherwise.

This brings us to why, on the prospective account, the constitutive aim of deliberation—deciding what to do—cannot be elevated to the status of a meta-end. For, it does not trigger deliberation, it does not enforce deliberation's conclusions, and it does not provide standards for deliberative success beyond finding some motivationally and epistemically viable option.

1.3.8 The process of deliberation

So the story goes: as an agent deliberates, she is searching for an option 1) that represents a course of activity that appears to be free from gaps, 2) whose represented course of activity does not appear to contradict the content of any other prospection currently guiding her behavior, and 3) whose represented course of activity is sufficiently stimulating.

It is important to clarify here, however, that this deliberative "search" does not involve merely a canvassing of pre-fabricated options. Rather, the agent must actively manufacture her options in imagination[40] by recombining, redescribing, specifying, and innovating the content of various options until she finds one that works. Recombination involves splicing together content from competing lines of action into a single option. Redescription involves reclassifying the activities involved in an option by attending to them in more detail, or in the light of other contextually relevant features (e.g., redescribing being on time to a meeting as being respectful to one's colleagues). Specification involves making a general plan more specific. And innovation involves inventing new concepts or using old concepts in new ways.

Deliberation then, is a matter of problem-solving. It is a matter of using these four tools—recombination, redescription, specification, and innovation—in order to craft an option free from any unsettling prospective content. Importantly, when an agent finds an option that works, no additional act of "will" is needed to settle her prospective set. Because she has manufactured an option that passes all bars for incorporation, it will be

[40] Imagination is the faculty with which we entertain non-actual states of affairs.

activated, and she will straight away return to motivated behavior, with the content of the successful option simply becoming the content of an activated prospection.

Also importantly, a deliberator can apply the processes of recombination, redescription, specification, and innovation to any prospection in her prospective set, not just the one that originally caused the motivational problem. Faced with the raging river, for instance, the hiker might not only deliberate about how to get to the campground, and whether to get to the campground, but also whether it is important to keep her shoes dry, whether it is cowardly to turn back, and whether she should abandon her ambition to be a professional wilderness guide. In other words, no particular goal in prospection is beyond the reach of imagination's revisionary processes. And, within the constraints of an agent's motivational dispositions, deliberation can involve the revision of goals as much as it involves the revision of means to those goals.

1.4 Deliberation and Deliberate Behavior

Because deliberation requires an agent to consciously attend to her options, she is typically aware, when deliberation concludes, of what she expects herself to do. After deliberating long and hard about whether to join her friends on a road trip, say, an agent is aware of, and has accepted via deliberation, a plan to forgo the road trip in order to finish painting the exterior of her house. Thus, she deliberately paints her house and deliberately declines the road trip offer.

However, we often consciously attend to and accept prospective content without deliberation. Imagine an agent who is being approached on the street by someone handing out fliers. She might, with no deliberation at all, form a plan to decline the advertiser's offer by saying, "No, thank you." Suppose that such exchanges always make

her feel a bit uncomfortable, so as the advertiser approaches, she is attending to the expected event in her imagination, dreading it in a small way. It seems here that the mere fact that she is attending to, and in the light of such attention accepting the goal of declining the flyer, is sufficient to qualify the behavior as deliberate. Thus, I propose as a definition of deliberate behavior any behavior that is guided by goals that 1) have been attended to by the agent, and that 2) have, in the light of that attention, remained practically engaged. Thus, we should expect all behavior that follows from deliberation to be deliberate, but other motivated behavior may be as well.

1.5 The Belief-Desire Dilemma

The account I have developed here suggests that we abandon the idea that motivated behavior is caused by a pair of mental states—belief and desire—and accept instead that they are caused by a single mental state, prospection. But, given my proposal that prospective content is governed by both epistemic and motivational constraints, we might wonder just how far we have diverged from the traditional belief-desire picture, particularly when it comes to accounting for deliberative phenomena and action.

To make this point clear, let us consider the way that belief-desire psychology factors into a persistent puzzle within ethics. The puzzle can be put as follows: The outcomes of practical deliberation must be either beliefs or desires. If they are desires, we can easily explain the motivating force of our practical judgments, but we must abandon the idea that practical judgments are the products of rational processes and subject to rational criticism. If, on the other hand, our practical judgments are beliefs, we

can preserve the idea that they are the products of rational processes and subject to rational criticism, but we surrender our ability to explain their causal efficacy.[41]

We might hope that abandoning the belief-desire psychology upon which this puzzle rests would help us resolve it in a clear way. But it does not, at least, not yet. For, it is possible to rephrase the problem in terms of epistemic and motivational constraints rather than beliefs and desires. Do the conclusions of practical deliberation place epistemic or motivational constraints on prospective content? For example, if I judge that the right thing for me to do in a given situation is to send my child to private school, is this judgment effective in shaping the content of my prospective set because it constitutes a belief that I should send my child to private school (i.e., a belief that I should possess $\phi_{\text{private-school}}$ as an activated goal), or because it actually activates that goal, or at least, makes me motivationally inclined to reject any plan that does not include $\phi_{\text{private-school}}$?

And here the dilemma repeats itself. As Pamela Hieronomy has argued, deciding it would be good to intend to ϕ is not the same as deciding to ϕ, just as deciding it would be good to believe p is not the same as deciding to believe p.[42] Hence, if our practical judgments are simply beliefs about which prospects we should and should not have as activated goals, it remains to be explained how they could play a causal role in actually activating or deactivating those goals. Further, given the cases of akrasia we face in our daily lives it seems that we have evidence suggesting the opposite. Our judgments about

[41] The belief-desire dilemma is related to the puzzle articulated by Michael Smith as "the moral problem." However, unlike the moral problem, the belief-desire dilemma does not emerge from any pretheoretic intuitions about whether an agent is, ceteris paribus, moved by her practical judgments, nor whether those judgments are judgments about "matters of fact." Rather the only pretheoretic intuitions we need are the following. Sometimes practical judgments move an agent to action, and in many, if not all of those cases, we believe those very judgments are subject to rational criticism.

[42] Hieronymi, Pamela. 2005.

which goals ought to be activated do not alone have the power to activate those goals. On the other hand, if our practical judgments just are activated goals, it is unclear in what sense these conclusions are the appropriate subjects of rational criticism. Further, it raises the question of why we form normative beliefs about which goals ought to be activated at all.

The following chapter is a response to this reformulated version of the belief-desire dilemma. I will show that the prospectivist account of motivation can provide the basis for an escape from this dilemma. Thus, although replacing belief-desire psychology with the prospective account of motivation does not alone resolve the puzzle, it provides the important first step.

Chapter Two

The Search for Rational Deliberation

To frame the investigation at the heart of this chapter, I open with a passage from Jay Wallace:

> Practical reason is the general human capacity for resolving, through reflection, the question of what one is to do…Our capacity for deliberative self-determination raises…questions about how deliberation can succeed in being practical in its issue. What do we need to assume — both about agents and about the processes of reasoning they engage in — to make sense of the fact that deliberative reflection can directly give rise to action? Can we do justice to this dimension of practical reason while preserving the idea that practical deliberation is genuinely a form of reasoning? [43]

This quote asks two questions of practical reason: First, in what sense is practical reason causally efficacious and, second, in what sense is it reasoning as opposed to just mental activity. In this chapter, I hope to demonstrate that a prospectivist motivational psychology successfully answers both these questions.

2.1 The Practicality of Practical Deliberation

If we take "action" in the above quote to be equivalent to "deliberate behavior" as defined in Chapter One, then prospectivism sufficiently addresses the first question above. Here is a summary of how it does so. In order to make sense of the fact that deliberative reflection is practical in its issue, we need to assume only the following about deliberating agents: First, motivated behavior is directly guided by prospections—

[43] Wallace, R. Jay. 2008. (1)

chains of goal-shaped expectations. Second, as long as each prospection in an agent's prospective set meets the three activation conditions—means-end contiguity, non-contradiction, and motivational sufficency—motivated behavior proceeds. Third, when motivated activity is interrupted because of a prospective failure—a scenario in which one (or more) of the prospections in an agent's prospective set fails to meet these conditions—an agent deliberates, using her faculty of imagination, to generate new, failure-free prospective content via recombination, redescription, specification, and innovation. When this process is complete, when she has successfully generated a failure-free prospective set, motivated activity is restored.

2.2 Defining "Action"

There is one caveat I want to mention in response to Wallace's use of the phrase 'directly give rise to action.' Again, if we take action to be equivalent to deliberate behavior then this phrase is unproblematic for my account. But some philosophers believe that action is a more substantive concept, that it picks out an even narrower class of human activity. If this is the case, then although I have answered the question of how deliberative processes are practical in their issue, I have not yet explained how they could give rise to action, in the substantive sense of the word. So here, I want to offer briefly an account of what I think the correct understanding of action is.

I believe setting action equivalent to deliberate behavior is appropriate because it situates it properly between one conception of action that is too broad—Hume's—and one conception that is too narrow. Criticizing Hume, Korsgaard writes, "Hume's view depends upon an inadequate conception of action. For Hume, an action is essentially

nothing more than a movement caused by a judgment or idea."[44] I agree with Korsgaard that Hume's view is obviously too broad. If Hume is right, then any motivated behavior, even the automatic, counts as action. But habitual and impulsive behaviors are among the prime candidates we want to exclude from the class of action, at least if we want, as most action theorists do, to exclude the majority of non-human animal activity from the class of action. This makes Hume's conception untenable.

However, some who want to maintain that practical reason plays a key role in distinguishing action from lesser forms of motivated activity overcompensate for the breadth of Hume's position, delivering a conception of action that is too narrow. I will use Velleman as the example of such overcompensation.

For Velleman, all action is aimed at and guided by one goal: self-understanding. One's actions contribute to one's self-understanding so long as they make sense of an agent's motivations within the agent's particular environment. For example, my decision not to gamble my last $100, even in the face of great odds for winning a very large pay-off, makes sense given how crucial that $100 dollars is to my survival, and how I value survival much more than being rich.[45] When we deliberate, Velleman believes, it is because we are uncertain which course of action will make sense given our goals and our environment, and we are trying to settle this question for ourselves. Of course, sometimes an agent may think she knows the answer to this question without deliberating, and in those cases, her action can be guided by the aim of self-understanding

[44] Korsgaard, Christine. 2009. (64)
[45] Velleman, David. 2000. This example is adapted from one Velleman uses in a critique of decision theory. His point is that the formal rationality prescribed by decision theory can sometimes be at odds with "rationality on the substantive dimension," (167) which includes viewing options in terms of concepts like "good enough," and that we should not forsake that latter for the former. (144-169)

without involving reflective deliberation. But whenever an agent's uncertainty about her own motives or her environment call into question her self-understanding, she must deliberate. Since the conclusion of her deliberation will be an action-guiding intention to do what makes the most sense, where deliberation has filled in the particular content of "what makes the most sense," action will regularly follow from deliberation.

Let me begin by noting that it is possible to create a Vellemanean deliberator from a prospective deliberator by adding to the prospective deliberator's motivational set a powerful, perhaps unconscious, desire to do whatever will make most sense of oneself. If there are in fact agents out there who are so motivated, and Velleman himself may be one such agent, then Velleman's account of deliberation is an accurate descriptive account of at least those deliberators. I do not claim, therefore, that deliberation as envisioned by Velleman is psychologically impossible. Rather, my claim is that his account is not general. It does not pick out the features common to all deliberators, but rather, the features common to some subset of deliberators with certain substantive inclinations.

On the prospectivist picture, self-understanding is useful to solving the problems one faces during deliberation. Thus, it might look as if all deliberators care ultimately about self-understanding. But, if self-understanding is a means to solving one's problems, then one does not need to care about it in itself in order to engage in the mental activities that bring it about, nor does it have any final authority or oversight with respect to the ordering of one's affairs. We should not assume that self-understanding is the meta-end of deliberation any more than we assume that being transported is the meta-end of driving, though of course, an agent might have an strong inclination to be transported.

Velleman's account of action is therefore too narrow because it ties action to a regulating goal only some agents possess. Remember Korsgaard's claim, "Human beings are condemned to choice and action…It is our plight: the simple inexorable fact of the human condition."[46] To make good on this claim, we must understand action broadly enough so that it follows from all complete deliberation, not just deliberation guided by an optional or idiosyncratic end. In contrast to Velleman's picture, prospectivism claims that we are all bound to run into prospective uncertainty at one point or another, regardless of whether we share any substantive aims. Thus, agency is tied to something we have in common with other humans, but this something is not a particular aim, a particular regulatory inclination, or even a particular problem, but rather the capacity to have and solve problems itself.

The question then becomes whether the processes that yield deliberate behavior (which I will henceforth use interchangeably with 'action') can be considered rational in any sense. As a starting point, I will examine two contemporary proposals on this question, one made by Candace Vogler and one by Bernard Williams. I will respond to the shortcomings in these views by offering my own account, and then I will contrast this account to a number of additional proposals.

2.3 The Rationality in Practical Deliberation

2.3.1 Vogler's instrumentalism

One popular view of practical reason is instrumentalism, the view that deliberation employs reason insofar as it involves the ability to grasp one's goals and reason about the means to those goals. One contemporary proponent of this view is

[46] Korsgaard, Christine. 2009. (2) Emphasis hers.

Candace Vogler, who has attempted to defend what she takes to be the key insight of instrumentalism, the proposition which, she believes, allows instrumentalism to maintain its status as the 'standard view.'

The proposition is this: practical reason is calculative, i.e., concerned with matching means to determinate ends. Further, calculative reason can tell the entire story of practical reason.[47] If this is the case, we should be able to formulate deliberative standards in calculative terms. In fact, Vogler insists that deliberation, even normative-flavored deliberation of ends, falls within her calculative conception of practical reason.[48] She argues for this as follows: Deliberation is constitutively aimed at something, namely, deciding what to do. If deliberation is constitutively aimed at something, it therefore has a calculative form, a form exposed with a 'useful-style' explanation: I deliberate about ends in order to decide what to do. And, finally, if deliberation has a calculative form, then it is not importantly different from motivated activity in general.[49] Which is to say, deliberation is more or less rational with respect to how well it delivers on its aim.

But here Vogler misses an important fact. Constitutive aims do not necessarily explain nor motivate the behavior which they are used to distinguish. To illustrate this point, imagine asking someone why she is playing chess, and having her respond, "I am playing chess in order to capture my opponent's king." In doing so, she has not actually answered your question. Rather, she has merely confirmed that she is indeed playing chess. To explain why she is playing chess, i.e., why she is sitting there trying to capture

---

[47] Vogler, Candace. 2002. (6, 149)
[48] She writes: "Although moral psychology is very important to work on ethics, calculative views of practical reason need not turn on any specific moral psychological commitments…the claim that practical reason is primarily calculative cannot be defeated by recourse to claims about practical reasoning [or] claims about moral psychology." Vogler, Candance. 2002. (24)
[49] Ibid. (168-170)

her opponent's king, she would have to offer some further end, an end potentially fixed by any one of the threefold goods Vogler takes practical reason to track (e.g., I am playing chess because it is pleasant, and/or useful for x, and/or befitting).

We can make the same point about the constitutive aim of deliberation. The claim that I am deliberating about my options in order to decide what to do is uninformative. We must appeal to some further end to explain why we are deliberating. However, in the case of deliberation of ends there is no fixed end to which an agent can appeal to explain her behavior. Rather, there are numerous conflicting ends, and the conflict among them explains the fact that she is deliberating. The ends that cause the problem continue to make their demands until she, the deliberator, manages to reconfigure them.

Importantly, this is not a point about whether an agent is consciously aware of the fact that she is trying to decide what to do. Rather, it is a point about whether that aim, consciously or otherwise, does any explanatory or motivational work. Prospectivism says no.[50] On its own, the aim of making a decision offers neither a source of motivation nor standards for success beyond what the practical problem that prompts deliberation already provides on its own. So, although we can point to what all deliberators have in common with a general teleological description—namely, "they aim at deciding what to do"—a proper portrayal of deliberative processes does not see them serving settled ends, meta or otherwise, but rather modifying prospective content in order to lead problematic ends toward reconciliation.

[50] This is not to claim that it is impossible for such an aim to do such work, just that the aim does not generally do this work.

Vogler seeks to defend the core insight of instrumentalism, which is, she insists, neutral with respect to moral psychology. But if prospectivism is true, the practical reason she offers has nothing to say regarding what makes one choice better than another. For, in the face of unsettled prospective content, calculative reason (at least not as she has portrayed it) says nothing regarding how best to reorganize the ends in question. Of course, after we have made a decision, we will be able to explain the resulting activity calculatively in light of the ends that ended up mattering. To return to the example from §1.3.7, we can say after I accept the Boston job that I did so at least in part because I value living where there is a robust public transportation system, and that living somewhere with such a transportation system is a conscious end of mine. But the value of the ends that might eventually factor into useful-style explanations are themselves in question during deliberation. This is what is special about the "much-touted"[51] practical point of view. Deliberation is in a class of its own; it cannot properly be considered just another motivated behavior.

So, against Vogler's attempt to defend the key instrumentalist insight, I seek to defend a key anti-instrumentalist insight, albeit without its standard Kantian framing. The insight is this. Deliberative processes do not serve antecedent ends; rather, they modify them. Accordingly, if there are any deliberative standards at all, these standards must be grounded somewhere besides antecedent ends.

But then, of course, we must wonder: Is there any principled way to make a decision, and, if so, where do the principles get their authority? If there are no principled ways to make a decision, it is difficult to see how deliberation could be considered a rational process at all. If all we are doing when we deliberate is randomly surfing options

[51] Ibid. (6)

in our imagination until we hit upon an option that satisfies the three prospective constraints, this seems like a decidedly arational process.

2.3.2 Williams and rational motivational change

In his article "Internal and External Reasons" Bernard Williams offers a response to the questions posed by Wallace above. Williams maintains that deliberative processes can impact the contents of an agent's subjective motivational set, S, because the mere entertainment of various propositions, in the service of generating a viable option, have motivational impacts on an agent. This accords nicely with the prospectivist picture of practical deliberation, which involves using imagination to redescribe, recombine, and specify the contents of various options until one meets all three activation conditions. In other words, in order to be successful, deliberation not only can, but must, produce a change in an agent's subjective motivational set; it must transition an agent from having an S currently unequipped to guide behavior to having an S that actually does guide behavior. Williams is right when he writes: "We should not…think of the contents of S, as statically given."[52]

Importantly, however, Williams wants to make an additional, more ambitious point about rational practical deliberation. He wants to maintain that at least some of the processes employed in deliberation count as rational processes or practical reasoning. Consider the following passage:

> (W1) …indeed the mere discovery that some course of action is the causal means to an end is not in itself a piece of practical reasoning. A clear example of practical reasoning is that leading to the conclusion that one has reason to ϕ because ϕ-ing would be the most convenient, economical, pleasant, etc. way of satisfying some element in *S*, and this of course is controlled by

[52] Williams, Bernard. 1979/2001. (81)

> other elements in *S*, if not necessarily in a very clear and determinate way.[53]

Although Williams seems confident that the former process is not an instance of reasoning and that the latter is, it is hard to know what Williams takes to distinguish these two cases. And since this is precisely the question I am interested in, I would hope for him to provide grounds for his distinction. Later, he writes,

> (W2) There is an essential indeterminacy in what can be counted a rational deliberative process. Practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion.[54]

Assuming he is right about the vagueness of the boundaries between rational and non-rational processes, this indeterminacy does not get him off the hook in terms of explaining why <u>in the paradigm cases</u> one process (the mere discovery of a means to an end) is not a rational process and why another (the concluding that one has a reason to ϕ because ϕ-ing is the most economical means to an end) is a rational process.

Perhaps vagueness lies in between, but what I hope to get from Williams is an account of the difference between what he has called a "clear case" of rational deliberative processes and those deliberative processes he thinks are clearly not rational. But in providing any such account, he must clarify what it is to be a rational process in a way that makes it clear that rational processes can have motivational impacts. For, if it turns out that only the non-rational deliberative processes are the ones capable of delivering the motivational changes required for successful deliberation, then Williams would be left with the conclusion that rational processes have no practical import, a conclusion he is eager to reject.

---

[53] Ibid. (80)
[54] Ibid. (86)

Because he doesn't provide an account of the distinction between rational and non-rational processes explicitly, we must extract it from his text. Consider the following two passages in which he muses on the different ways an agent can deliberate.

> (W3) …there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in S can be combined, e.g. by time-ordering; where there is some irresoluble conflict among the elements of S, considering which one attaches most weight to…; or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening granted that one wants entertainment.[55]

> (W4) [An agent] may think he has reason to promote some development because he has not exercised his imagination enough about what it would be like if it came about. In his unaided deliberative reason, or encouraged by the persuasions of others, he may come to have some more concrete sense of what would be involved, and lose his desire for it, just as, positively, the imagination can create new possibilities and new desires.[56]

The point that Williams ends on in W4—that imagination can create new possibilities and new desires, and destroy old possibilities and old desires—is entailed by prospectivism, and is also what gives credibility to Williams' assertion that the contents of one's subjective motivational set are never statically given, and can be altered via imagination-driven deliberation. But this doesn't yet say anything to the question of whether the processes by which imagination creates and destroys desires are rational or non-rational, nor does it say how reason is involved one way or another.

One possibility is that rational practical deliberation involves not just the creation and destruction of possibilities and desires, but also a process consisting of an agent's judging what he himself has reason to do. This suggestion is supported by passages W1 and W4. In both passages, an agent's assessment of his reasons for action seems to be a natural sub-process of deliberation. In W4, for instance, Williams portrays a deliberative

[55] Ibid. (80)
[56] Ibid. (81)

situation in which an agent's judgment regarding what he has reason to do hinges on whether his imagination has yet to destroy a certain desire. But we might wonder, how could this be so? Here is one set of circumstances under which this could be so:

(a) The truth of internal reasons statements hinges on facts about what an agent desires.

(b) A deliberating agent knows that the truth of internal reasons statements hinges on facts about what he desires.

(c) A deliberating agent is attempting to make a true judgment regarding which internal reasons he has.

(d) When an agent's coming "to have a more concrete sense of what would be involved" in some ϕ, causes him to "lose his desire" to ϕ, he also thereby changes his judgment about having a reason to ϕ.

This schema might work as a generalization of some instances of deliberation. Imagine a case in which an agent thinks he has reason to cheat on an exam, but on reflection he realizes that doing so would violate principles he respects and wants to endorse. He also reflects on the consequences of not cheating and realizes that even the worst-case scenario—failing the exam—is not all that bad. Having a more concrete understanding of his options thus completely kills the desire he once had to cheat, and, in virtue of this fact, allows him to correctly judge that he has no (internal) reason to cheat.

However, even if this offers an accurate description of how people sometimes deliberate, and even though it involves an agent's reflection upon what he has reason to do, it is flawed as an example of a rational process of deliberation for two reasons.

First, in a point made by Michael Smith, it is not the case that in all instances in which imagination alters our desires we are inclined to say that the resulting desires have

been rationally produced.[57] For instance, if an agent is scheduled for surgery to remove his gangrenous foot, he might allow himself to dwell in imagination on the pain of the surgery and the difficulties of life as an amputee. If these imaginative dwellings evoke strong aversions, they might thereby shift the contents of his prospective set so that he decides not to show up for surgery. But it seems to run contrary to intuition to claim that such a decision has been rationally arrived at, even if the agent concluded in response to his changing desires that he had a strong reason to avoid surgery. And notice, the problem here is not that he misimagined anything. Surgery is painful, and life as an amputee is hard. It seems appropriate that dwelling on such an imagined future evokes aversion. It is only if the aversion evoked is enough to sway the agent to avoid surgery, that his decision seems irrational. This raises the question of whether any motivational impacts that are prompted by imaginative specification, redescription, or recombination count as rational changes.

Second, if changes in an agent's motivational set are what make the causal difference in deliberation, as prospectivism entails, and an agent's judgments of what he has reason to do simply respond to changes in the motivational set as they occur, but do not actually cause any changes themselves, then even if such reason judgments are part of deliberation, they are causally inert. They do not actually help an agent determine what he is to do, even if they help an agent decide what he has reason to do.

Taken together, these objections make me suspect that Williams' must have something else in mind. We cannot make the imaginative processes at work in effective deliberation rational simply by laying atop these processes an agent's (causally

[57] Smith, Michael. 1994. (156-158)

inefficacious) process of judging what he has reason to do. So what else could Williams have in mind? Evidence may be found in this passage:

> (W5) I take it that, insofar as there are determinately recognizable needs, there can be an agent who lacks any interest in getting what he indeed needs…If an agent really is uninterested in pursuing what he needs; and this is not the product of a false belief; and he could not reach any such motive from motives he has by the kinds of deliberative processes we have discussed; then I think we do have to say that in the internal sense he indeed has no reason to pursue these things.[58]

Implied in this passage is that an agent has an internal reason to ϕ if an agent can arrive at a state of being motivated to ϕ by employing the kinds of deliberative processes described in W1, W3, and W4 to get from previous motivational states to the state of being motivated by ϕ. So we might think that Williams is suggesting that these processes that change desires are at the same time rational processes because in changing desires, they simultaneously generate reasons.

But this gets us nowhere. If Williams is suggesting that <u>all</u> of the deliberative processes discussed in his text yield reasons (thus making them rational processes), then he is subject to the objection illustrated by the surgery example above—sometimes the imaginative processes involved in deliberation yield changes in motivation that seem to explicitly contradict what even an agent himself might judge he has reason to do. Or, as Smith puts it in a similar objection, "I see no way in which the effects of compulsions, addictions, emotional disturbances, and the like could be precluded by [Williams' account of rational deliberation]."[59]

If Williams believes that only <u>some</u> of the desire-changing processes he has discussed yield reasons (thus making them rational deliberative processes) and some do

[58] Williams, Bernard. 1979/2001. (81)
[59] Smith, Michael. 1994. (158)

not, then he still has not provided the criteria we would need for distinguishing between the rational and non-rational processes.

2.4 Diagnostic Reason

2.4.1 Diagnosing a prospective problem

I suggest the following understanding of practical reason. Reason is involved in practical deliberation insofar as it helps an agent locate, diagnose, and articulate the causes of a given prospective uncertainty, so that her deliberative search for a resolution becomes progressively less random and more intelligent. I will call this diagnostic reason. When employed during practical deliberation, diagnostic reason is practical in two senses: the problem it aims to diagnose is a practical problem, and the solution it helps deliver is action itself.

There are three primary benefits to viewing the role of reason in this way. First, it will help explain why sometimes reason appears to deliver deliberative conclusions, but sometimes it does not. For, sometimes a proper diagnosis of a problem is all that is needed to answer the question of what is to be done. This is the case when one already has a trusted method for resolving the problem on hand, so that once one realizes that she is facing that problem, she can easily identify and enact the solution. Once I realize that my car is pulling to the left because of a flat tire, I know exactly how to prevent my car from pulling to the left. But if I realize that my car is pulling to the left because of alignment issues, which I can recognize but do not understand in depth, I do not yet know how to make my car stop pulling left.

Second, it can explain why in some cases of practical deliberation, the choices we settle on are provisional. For, if our diagnostic reason locates a problem we do not

understand well enough to solve, we have no choice but to experiment, to adopt an attitude of, "I'll try this and see what happens." But even in these cases, if our diagnostic reason has served us well, our experimentation is not entirely random. I will not try to fix my alignment problem by releasing rodents into the undercarriage of my car, or by removing the windshield, or by driving the car into a brick wall. Diagnostic reason can get me as far as knowing which experiments might work, or at least which experiments might yield useful feedback.

Third, diagnostic reason can account for the fact that practical rationality is honed over time, and that it is more reliable as a deliberative tool in situations that are familiar. In diagnosing a problem with a car, for example, someone who has seen and solved many problems with such cars will likely be more effective in solving the problem with the car.

In response to this third point, some might argue that it is not the capacity for reason that is getting honed in these situations, but rather one's store of background knowledge. But, at least when it comes to diagnostic reason, the capacity to engage in a fruitful diagnosis of a given problem does not involve "accessing" background knowledge via reasoning processes, processes that then check such background knowledge for relevant information. If that were the case, an agent would need to check her entire store of background knowledge every time she encountered a problem. Our minds are more efficient than this. The lessons of experience are built right in; our capacity for diagnostic reason is constituted by our ability to recognize problems as familiar types, and to see in new problems relevant analogies to other problems we have encountered.

Empirical psychologists have studied this phenomenon. Michelene Chi, a psychologist at the University of Pittsburgh, writes, "It is well known by now that the quality of problem representation influences the ease with which a problem can be solved."[60] In one study, Chi and her fellow researches documented differences in how physics experts and physics novices categorized the problems in a given problem set. The novices categorized problems according to "surface features" such as into groups of problems about "rotational things" or "blocks on inclined planes." In contrast, experts were able to quickly group the problems in accordance with the major physics principle that would allow for their solution—for example, into groups of problems that "can be solved with Newton's second law" and "can be solved with the Law of Conservation of Energy"—even though the surface features of the problems were very different.[61] The researchers concluded that the initial identification of a problem schema "restricts search for a particular solution to a small range of possible operations."[62]

Of course, the examples from Chi's study deal with fixed-end reasoning. In these cases, there is a particular, unchallenged goal—finding the answer to the physics problem—and diagnostic reason helps the physicists figure out which method to apply to achieve this goal. But much practical deliberation, I have argued, involves cases in which our goals themselves are in question because of a particular conflict. In such cases, diagnostic reason is employed to help us locate and clearly understand which ends and means need to be changed, how much they need to be changed in order to resolve the problem, which ends and means can be changed, and what other problems may arise when such changes are implemented. By giving us a clearer understanding of the

[60] Chi, M. T. H., Feltovich, P., & Glaser, R. 1981. (122)
[61] Ibid. (125)
[62] Ibid. (150)

problems we face, our diagnostic reason assists our imagination in its search for acceptable prospective content. Diagnostic reason allows us to rule out instantaneously whole classes of options, and it helps us locate those features of the remaining options that point us toward prospective revision that will work.

Understood this way, then, diagnostic reason does not itself activate or deactivate prospective content. Its conclusions are not propositions of the form "it is right to do x," but rather intelligent search parameters along the lines of "look here for a potential path forward" and "doing a will lead to b, and b will be hard to live with." Diagnostic reason guides us toward content that will, in virtue of meeting the three activation conditions for prospection, activate itself. An apt analogy for deliberation, then, is someone who is actively trying to pick a lock using various tools. Diagnostic reason will help the agent understand the shape, size, and pressure that needs to be applied in order to open the lock, but the "key" will be complete when and because it opens the door. Similarly, an agent's choice is made when and because a particular imagined option suffices to meet the activation conditions.

Importantly, if this successful option has been forged in the light of a sophisticated and accurate understanding of the agent's problem, then reason has played a significant role in delivering practical choice. When deliberation involves such processes it is rightfully called rational. The inherent motivating properties of the successful prospective content discovered through this process will make the choice effective, and guide an agent's actions, but the content itself is shaped heavily by an agent's understanding of her problem, an understanding that is subject to rational criticism.

Soon, I will address the precise ways in which problem schemas are subject to rational criticism.

2.4.2 Not reason enough

The picture of reason I have offered here cannot be reduced to simple means-end reasoning. For, reason here does not simply serve an agent's activated goals and plans. Rather, it assists deliberating agents as they creatively manufacture new goals and plans, and thereby new motivations. Still, some may see this as a decidedly Humean picture. Reason, it seems, provides no motivational force. Its causal efficacy, rather, lies in its ability to assist an agent in identifying and authoring options that harness the force of motivational and epistemic dispositions over which she has no direct control.

This is quite a departure from the more full-blooded Kantian versions of practical reason that writers like Korsgaard, Wallace, and Velleman seek to defend. On such pictures, reason is much more than a tool for diagnosing our motivational problems. It is an independent spring of action. It delivers both a definitive verdict on what should be done, and also (in normal functioning adult agents) the causal power to make that verdict a reality.

These theorists might accuse me of defining reason in a manner that begs Wallace's first question, the question of how practical reason can be causally efficacious. In my original answer to that question, I claimed that practical deliberation is effective because it is, by definition, the process that restores motivated activity when motivated activity has been interrupted. But deliberation is not itself a rational process, but rather an imaginative process influenced by diagnostic reason. Hence, I have not shown that practical reason is itself effective, but rather that practical reason is part of an effective

deliberative process. We might worry that this is quite like claiming that some sugary cereal is part of a complete breakfast, when all of the "completeness" is being achieved by the other nutrient-rich components of the meal.

In response, I first want to highlight the fact that diagnostic reason is neither strictly theoretical nor practical. It is a capacity to develop problem schemas in the face of uncertainty, schemas that then direct a search toward new, unproblematic content. This capacity can be applied either to descriptive uncertainty (uncertainty about the facts) or normative uncertainty (uncertainty about what to do). Of course, insofar as problem schemas are representations of problems, they are themselves descriptive, and as such, they can be more or less accurate. This might spur the intuition that the work of diagnostic reason, like means-end reasoning, is theoretical work that just happens to have, because of the context in which it is employed, practical consequences.

However, it is important to remember that even though problem schemas can be more or less accurate, the primary aim of these schemas during deliberation is not accuracy, but usefulness. Returning to the case of the physics problems above, the novices who grouped the problems according to surface features did not have inaccurate representations of the problems, they just had less useful representations of the problems. Most philosophers have, rightly I believe, rejected the pragmatic notion that theoretical reason, and the descriptive beliefs it yields, aims at usefulness as opposed to accuracy. Hence, if diagnostic reason aims to deliver useful problem schemas, it is not itself an instance of purely theoretical reason.

But then we may wonder, is diagnostic reason's involvement in deliberation sufficient to make the conclusions of deliberation susceptible to rational criticism? I

believe it is. Insofar as an agent's diagnostic reason yields an understanding of the problem she is facing, and insofar as this understanding restricts the options she considers during deliberation and thereby restricts her eventual behavior, we can conclude that her understanding of the situation seriously influences her actions. This understanding is subject to rational criticism—it can be criticized both on grounds of inaccuracy and on grounds that some more useful understanding was overlooked—and, derivatively, any actions under its guidance are also subject to rational criticism. Another way of putting the point is this: when we criticize an agent's actions as irrational we are in fact aiming our criticism at what we believe to be an agent's faulty understanding of her situation.

2.4.4 The objects of normative uncertainty

Diagnostic reasoning can be deployed in practical deliberation for two conceptually distinct tasks: fixed-end reasoning and open-ended reasoning. Fixed-end reasoning, i.e., means-end reasoning, is characterized by an agent's endorsement (or, perhaps, more minimally, acceptance) of all activated goals in prospection. Because circumstances do not create a demand for the revision of goals, diagnostic reason is set only the task of specifying the means to activated goals. Open-ended reasoning, however, is required when two or more activated goals are in conflict, or when one activated goal is not realizable given environmental circumstances.

The problems that call for open-ended reasoning are marked by the need to adjust one's goals, and solving these problems inherently involves questioning the import of preserving certain goals as is, versus revising or abandoning them. Thus, when we are struggling to decide, in the face of conflict, which goals to endorse and continue to invest resources in, normative theory becomes relevant as a problem-solving tool.

Remember, on the prospectivist picture, the true objects of practical uncertainty, i.e., the what in what we are uncertain about when deliberating, is our prospective set. In short, we are uncertain regarding what to expect from ourselves and from the environment with which we are interacting. But because conflicts among goals and the environment can cause open-ended problems, prospective uncertainty gives rise to normative uncertainty, in the following way.

Suppose you have internalized two norms: do not eat animal products and be a gracious house guest. These norms have been conditioned as immediate evaluative rejections of certain prospective contents. Now, when your host offers you for breakfast a plate of eggs and bacon and a glass of milk, your norm against consuming animal products rejects this future and causes you to search for a replacement. But the nearest replacement, declining the food, also is rejected, by a different norm. Part of diagnostic reason's job when conflicting norms cause problems for prospective content is to ask about the possibility and consequences of flouting either norm.

A shorthand way to describe this is as a process of evaluating the import of the norms in question. Further, as Dewey would remind us, in almost any case of prospective conflict, as we delve further into the diagnostic work, looking for possible ways we might revise our goals, we eventually are forced to ask questions of import. Recall the hiker from §1.3. Faced with the raging river, she might not only ask how to get to her campsite by nightfall, but also whether it is important to get to her campsite by nightfall, whether it is important to keep her shoes dry, whether giving up and turning back is cowardly, and whether it is important to avoid being cowardly. The line between fixed-end reasoning and open-ended reasoning is hazy.

Assessments of import will involve attending to the phenomenology and content of one's occurrent desires, attending to the expected consequences of abandoning or preserving particular goals, and considering how given goals and values function to harness, effectively and reliably, motivation in unproblematic ways. Thus, a sophisticated understanding of the problems we face in deliberation will involve theories about which of our goals and norms are important and why. In other words, a sophisticated understanding of our problems will involve normative theory.

2.5 Locating Diagnostic Reason

The way we understand our problems influences how robust and satisfactory our solutions to those problems will be. To make this point clear we can consider the tragic case of Mary Griffith, who, upon learning her son Bobby was gay, convinced him to pray to God to be cured and to turn to church activities for solace and guidance. Bobby heeded her advice, and the pressure from his community, but these actions only led him to loathe himself more, until he eventually committed suicide. Mary has since left her church and become an active advocate for gay and lesbian youth.[63]

Both Bobby and Mary misunderstood the contours of their respective problems. Some of this misunderstanding involved mistaken means-end beliefs. For instance, Mary thought that her son could change his sexual orientation through prayer. But they also had inadequate understandings of whether and why it was important to "cure" Bobby's gayness. They drastically overestimated how hard it would be to live a life of open embracement of Bobby's sexuality and drastically underestimated how hard it would be for Bobby to live the life recommended by his family and church. Mary put too much

[63] Aarons, Leroy. 1995.

stock in the importance of how her family would be perceived, and not enough in the psychological health of her child. Further, the whole community failed to take note of the contradictory evidence bearing on their beliefs about the afterlife consequences Bobby would face if he did not repent.

The following lesson can be gleaned from the above example. Diagnostic reason plays a central role in guiding agents' behavior. Therefore it is legitimately practical reason. It will be informative now to contrast this conception of practical reason—as the capacity to accurately and effectively diagnose one's prospective problems—with other conceptions of practical reason on offer.

2.5.1 Velleman's self-understanding

Velleman's approach to solving the belief-desire dilemma introduced at the end of Chapter One is to posit the existence of a strong, mostly subconscious inclination toward self-understanding. This grounds a solution to the belief-desire dilemma in the following way. Faced with a variety of options about what she might do, an agent forms a belief about which of the options would make the most sense of herself given her motivations and the environment, and then her inclination toward self-understanding adds its motivational force to that option.[64] The result is an intention, a mental state that has the same direction of fit as belief—it aims to accurately represent the future—but the same direction of guidance as desire—it guides an agent's actions so as to make itself true.

[64] Velleman, J. David. 2000. (22, 24, 141) In these passages, Velleman writes of the aim of self-understanding "adding its motivational force" to first-order desires in order to tip the balance in its favor. This makes it seem that Velleman takes the aim of self-understanding to interact hydraulically with first-order desires. He moves away from this position in more recent work, clarifying that the aim of self-understanding governs first-order desires, but does not compete with them. (cf. Velleman, J. David. 2008.)

Intentions are a special kind of belief. They are self-fulfilling predictions. In short, we fix the contents of our belief regarding what we will do by deciding what will make the most sense, and then we do what we believe we will do because we could not make sense of ourselves otherwise.

Contrast this picture with prospectivism, which takes doing what we believe we will do as basic to all motivated behavior, insofar as our settled prospections are self-fulfilling expectations. On this picture, intentions are viewed as a subset of settled prospections, all of which are causally efficacious. Thus, we do not need to posit any particular agency-bestowing inclination to explain the possibility of self-fulfilling expectations. Such expectations are a possibility independent from any particular aim or inclination.

Once Velleman posits that having an inclination toward self-understanding is what enables a certain subclass of beliefs (with their mind-to-world direction of fit) to acquire the same direction of guidance as desires, then self-understanding is also available to stand in as the motivating and organizing force (i.e., meta-end) behind practical reflection. But since prospectivism does not need to posit a given aim to explain the efficacy of expectations—it takes such efficacy as basic—nor to explain the advent of deliberation—it posits prospective uncertainty here instead—then injecting self-understanding as the rationality-bestowing property of deliberation is gratuitous.

Self-understanding is, of course, a critical component of diagnostic reasoning as I envision it. For, the fuller and more robust an agent's understanding of herself is, the more accurate and empowering her understanding of her problem will be. Thus, most deliberating agents are in fact working to increase their self-understanding. However, the

regular pursuit of reflective self-understanding is compatible with either possibility: that self-understanding is a necessary means to practical problem-solving or that self-understanding is the organizing goal of practical deliberation.

In his more recent work, Velleman encourages us to read his claim that self-understanding is the aim of reflection cautiously. He writes, "[L]ike method actors, we never stop relying on our first-order impulses to fund our activities; we merely shape those activities into the most understandable enactment of our impulses, all considered."[65] He wants to ensure that we not interpret the aim of self-understanding as competing with other aims, but rather seizing upon their malleability to shape those aims in understandable ways. The problem for Velleman is that he cannot explain why an agent would bother shaping activities into an understandable enactment of her impulses, without positing the existence of a tendency toward self-understanding. Here he pounds the table. We all have this tendency, and having this tendency just is what makes us rational.

The prospectivist agrees with the first half of Velleman's claim—we never stop relying on our first-order impulses to fund our activities—but disagrees with the second half. In contrast, the prospectivist maintains that we shape our activities only when and because the impulses are incapable of doing so on their own. In the process of refashioning our ends, self-understanding helps us get things back on track. But this process does not necessarily aim for, nor result in, the most understandable enactment of our impulses.

[65] Velleman, J. David. 2008. (432-433)

Velleman writes, "the drive toward self-understanding [exerts] a fairly minor, modulating role in our practical affairs."[66] The prospectivist theory does not need to posit a drive toward self-understanding at all. It explains why we are disposed to modulate our impulses in light of our self-understanding by taking self-understanding to play a crucial role in problem-solving and problem-preventing. It trades the claim that we are all regulated by an inclination toward self-understanding with the claim that we all have problems and the reflective capacities to help ourselves tackle those problems.

2.5.2 Smith's systematic justification

Michael Smith envisions practical reason as the process that allows us to form an evaluative belief "about what would be desired if we were fully rational."[67] The ideal of full rationality, for Smith, involves being such that all of our desires are systematically justifiable. Thus, deliberation is rational insofar as it delivers judgment on which desires are parts of an agent's systematically justifiable set of desires and which are not. He writes, "…though the imagination can indeed produce new and destroy old desires via vivid presentations of the facts, its operations are not guaranteed to produce and destroy desires that would themselves be sanctioned in an attempt at systematic justification,"[68] and "…by far the most important way in which we create new and destroy old underived desires when we deliberate is by trying to find out whether our desires are systematically justifiable."[69] Thus, for Smith, practical reason is a reflective process with a fixed and causally efficacious meta-end: systematic justification of one's desires.

---

[66] Velleman, J. David. 2008. (433)
[67] Smith, Micheal. 1994. (160)
[68] Smith, Micheal. 1994. (161)
[69] Smith, Micheal. 1994. (158) Emphasis his.

My response to Smith is similar to my response to Velleman. The aim of systematic justification of one's active goals is helpful as a means to overcoming the problems diagnostic reason is employed to solve, which is why it seems that deliberation often involves trying to identify systematizing relations between disparate goals. But such systemization is not the end of practical reason. Practical reason aims to find only enough systemization as is required to solve the problem. When it comes to systematic justification, diagnostic reason satisfices.

Which is to say, we do not need to appeal to the ideal of complete systematic justification to account for the fact that the imaginative processes in deliberation are rational processes. If the goal of diagnostic reason is only to locate and understand the forces (both internal and external) that have led to uncertainty so that an agent's search for a resolution is more effective, this will be enough to make it a rational process. This allows us to treat the ideal of complete systematic justification of desires as an optional ideal for rational deliberators.[70] And if the ideal is optional, so are any of the specific imperatives it issues.[71]

---

[70] I would like to register here my skepticism that 1) complete systematization of desires is possible, 2) even if it is possible, it is desirable, and 3) even if it is possible and desirable, it can be meaningfully executed in absence of particular prospective problem.

[71] I must admit here that my argumentation on this topic is hindered by my lack of clarity regarding Smith's understanding of systematically justified desires. At one point (158-161) he alludes to a Rawlsian process of reflective equilibrium to explain how systematic justification might be achieved. In general, I find reflective equilibrium to be an attractive method of conceptual analysis and theory-modification. But I am not sure what role Smith envisions reflective equilibrium playing in the face of prospective and normative uncertainty, i.e., in the face a given practical choice. He writes, "…we might find that our specific value judgments would be more satisfyingly justified and explained by seeing them as all falling under a more general principle...[and this] may itself cause us to have a new, underived desire for that more general thing." (160) Here the concepts of 'justified' and 'explained,' appear to go hand-in-hand. But, consider that an agent could be in the following position: she could realize that two of her value judgments could be unified under a single egostic principle and so adopting that principle would make her set of profiles more explicable but perhaps no more justifiable than it previously was. What would the ideal of systematic justification recommend with regard to adopting a more

2.5.3 Millgram's practical induction

Diagnostic reasoning, as I have been describing it, is a conscious, reflective, deliberative process that functions only in the face of an agent's actual uncertainty regarding what she will do. However, in order to do its job properly, diagnostic reasoning must be supported by continuous and automatic belief- and value-updating processes which may also fall under the umbrella of rational processes.

Prospectivism accepts that adjustments in ones beliefs, values, and goals can happen without conscious attention so long as those adjustments do not cause a prospective problem for an agent. Hearing a noise outside, I form the belief that someone is mowing the lawn without consciously attending to that belief. However, if the noise grows louder and louder and my house starts shaking, then I will be uncertain about the cause of the sound, and will need to do some conscious diagnostic work.

Elijah Millgram has argued that just as we form beliefs through processes of theoretical induction, so too do we form values and goals through what he calls "practical induction."[72] This conception of practical reason dovetails nicely with mine. Practical induction can be viewed as a background process critical to the proper functioning of diagnostic reasoning.

---

general egoistic desire in this case? Indeed, Smith suggests later that a particular egoistic principle—namely, the principle that one should gain wealth no matter what the cost to others—is indefensible because "fully rational creatures would want no such thing" (195). But of course we cannot make that judgment unless we assume that we have access to what fully rational creatures want. Smith writes that the "only court of appeal there is for claims about what we have normative reason to do" is "[people's] stock of wisdom about such matters against which each person's opinions should be tested" (195-196). This makes the systematization of desires an inherently public activity. But going public does not resolve the fundamental question if the public is not itself comprised of rational individuals. Whether we are trying to systematize only our own desires or the desires of everyone, we face the same question: does explanation via systematization really amount to justification?

[72] Millgram, Elijah. 1997. (43-51)

When an agent asks herself during conscious deliberation whether adopting a particular course of action will solve her prospective problem, she will need to consult a variety of values she has built up through past episodes of unconscious practical induction, values that now bear on her choice. Consider, for example, an agent attempting to decide whether to go back to school for a higher degree. Perhaps she has learned over time, for example, that a certain amount of disposable income makes her life much less stressful. Or, through a series of frustrating experiences, she has come to care more than she thought she would about prestige and social status. Or, her previous experiences in school have taught her that she tends to become frustrated and impatient in the classroom, preferring instead "real-world" situations. These are values and preferences she has built up through practical induction prior to a moment of reflection that now become relevant to her resolving a particular uncertainty.

She may also draw upon consciously directed episodes of practical induction during deliberation to form new values. For example, she may direct herself to attend to the experiences of all those she knows with higher degrees. Having never considered these disparate lives (say, of her father, of her high school friend, of her doctor) in conjunction, she now takes them as a collection of evidence that bears on the question of whether she should attain a higher degree.

To be clear, I see induction—both theoretical and practical—as supportive of diagnostic reason, but not constituting diagnostic reason. In order to home in on a useful problem schema, and to assist in the construction of a resolution to a particular deliberative uncertainty, an agent will need to make inferences. Practical induction will be incredibly useful in this regard.

2.5.4 Richardson's specificationism

Henry Richardson's conception of deliberation also dovetails nicely with the prospectivist account of motivation. Richardson has effectively defended the idea that we can, and regularly do deliberate about our ends, and has argued persuasively against maximization accounts of practical reason.[73] To account for how processes of deliberate revisions of ends are subject to rational pressures, he introduces the notion of specification.[74]

Despite our shared perspectives on the plight and needs of deliberating agents, it is worth noting the difference in emphasis between our two projects. Richardson is interested in laying out the conditions under which a given revision of ends, i.e., a given move from E to E^* could count as a rational revision. Then, by demonstrating that such revisions can apply not only to particular ends, but to ultimate ends as well, he shows that rational revision of ultimate ends is possible. The rationality of a given specification is, if I read Richardson correctly, determined by the relationship between the content of a given antecedent end and the resulting revision to that end.[75]

In contrast, my interest here lies not in the possibility of a rational relationship in content between one end and its revision, but rather in the nature of the mental capacities that can deliver revisions to ends in the first place. There may very well be a tight connection between these two projects, but I will not explore it here.

---

[73] Richardson, Henry. 1997. (See, in particular, 89-158)
[74] He uses this term more broadly then just the making general plans more specific. It also involves, for instance, getting specific about under exactly which circumstances a given end will be adequately satisfied, and so forth.
[75] Richardson, Henry. 1997. (69-77)

Chapter Three

Normative Authority and Deliberation

Humans need guidance, and they need it because they are reflective beings of a particular type. A handful of contemporary philosophers have emphasized that normative inquiry, including ethical inquiry, should not proceed without an awareness of the reflective conditions that give rise to the need for guidance. Call this the guidance-first approach to normative theory. I endorse such an approach, and in this chapter, I hope to defend it in two steps. First, I will explore and argue against Nomy Arpaly's explicit rejection of guidance-first approaches to normative inquiry. Second, I will suggest a prospectivist rendering of the type of guidance needed by practical deliberators. This will stand as a corrective against the flawed accounts of deliberation at the heart of other guidance-first approaches and will thereby make the approach more compelling.

3.1 A Guidance-First Approach

In The Sources of Normativity, Korsgaard frames normative inquiry as follows,

> When you want to know what a philosopher's theory of normativity is, you must place yourself in the position of an agent on whom morality is making a difficult claim. You then ask the philosopher: must I really do this? Why must I do it? And his answer is his answer to the normative question."[76]

Korsgaard then articulates three criteria that any answer to the normative question must meet. First, the answer must actually convince the sincere deliberator that she is bound in

[76] Korsgaard, Christine. 1996. (16) Emphasis hers.

the particular sense in question, or as Korsgaard puts it, the answer needed "is really the first-person answer, the one that satisfies us when we ourselves ask the question."[77] In a related point, the answer must be "transparent." Finding out why I think I must do something must not thereby convince me that I do not need to do it. Third and finally, Korsgaard argues, the answer to the normative question must appeal "in a deep way" to our sense of our identity. She argues that since morality sometimes demands hard things of us, including in some situations the sacrificing of our own lives, the answer to the normative question must show that sometimes "doing the wrong thing is as bad as or worse than death,"[78] and since a loss of identity is often just as bad or worse than death, she believes that the answer to the moral question must issue from a sense of who we are.

By demanding that normative philosophy must speak directly to the sincere practical deliberator, Korsgaard has implicitly staked out a meta-philosophical position—a position regarding the aim of philosophical inquiry itself. She believes that, at root, all genuinely normative inquiry arises from, and answers to, sincere deliberators. I am inclined to agree with Korsgaard on this point. But it is puzzling that Korsgaard restricts her framing of the inquiry to an agent struggling with morality's claims rather than an agent struggling with the authority of prospective content generally speaking.[79] Consider the fact that moral norms are not the only norms that require difficult sacrifices. Fashion can require us to wear uncomfortable and potentially dangerous shoes. Social mores can require us to tolerate nearly intolerable people. Prudence can require us to sit at our computers on a lovely afternoon. Given this, we could pose a question very similar to Korsgaard's question using any of these sets of norms as a starting point: Place yourself

[77] Ibid. (17)
[78] Ibid. (17)
[79] Bernard Williams briefly notes this objection in his reply to Korsgaard. 1996. (210)

in the position of an agent on whom a particular norm or set of norms (fashion, etiquette, prudence, aesthetics, law, parenthood, morality, etc.) is making a difficult claim. Then ask the philosopher: What must I do now? And why must I do it? His answer is the answer to the normative question.

Below, I will defend the meta-philosophical claim that normative inquiry cannot be cleaved from the needs of genuine deliberators, i.e., that normative inquiry must answer to the agent asking the normative question. Further, I will defend the position that normative ethics must center on a distinctive, collaborative project that understands a) the dynamic role norms play in guiding motivated behavior and b) the sources of legitimate challenge to normative authority. I will call the general position that normative claims and theories must answer to deliberating agents the guidance-first approach to normativity.

3.2 Arpaly's Rationality

Nomy Arpaly has explicitly rejected the idea that normative inquiry must cater to deliberating agents. In Unprincipled Virtue, she argues that agents can be rational without deliberating and irrational even when they deliberate and act in accordance with what they judge to be the most reasonable option. Because deliberators are fallible in this way, Arpaly does not believe that a given account of rationality should be constrained by the requirement that it have the power to convince genuine deliberators of its authority. And yet, at the same time, she believes that such an account, once in hand, will have authority over deliberating agents whether they recognize the authority or not. That is,

these agents will be committing mistakes if they fail to meet the criteria for rationality.[80] She writes:

> There are at least two ways to think of theories of rationality. One way is to see the ideal theory of rationality as providing us with a manual of sorts: follow these instructions, and you will always make rational decisions or you will, at least, know whether or not you are acting rationally at the moment. Another way is to see theorizing about rationality as aiming, not at providing us with a manual, but at providing us with a third-person theory—a theory that tells us when people act rationally and when they do not, so that given full information about a person's circumstances, beliefs, and motives, one would be able to tell how rational or irrational said person would be in performing an action. These two tasks…are more different than they may look, and conceptual blunders may ensue if one does not always know which of them one pursues.[81]

Because I seek to defend a guidance-first approach to normative inquiry, and because Arpaly's conception of a manual for rational behavior seems reminiscent of such an approach, I believe something relevant can be learned from exploring Arpaly's treatment of this distinction. Arpaly explicitly aims to contribute to the latter project, and she expresses doubt that there is any fruitful relationship between the two projects. This doubt takes different forms. Sometimes it seems she is merely dismissing the possibility that accounts of rationality can be used as manuals to rational behavior. Other times it appears that she is questioning whether a satisfactory manual to rational behavior is possible at all, and thereby implicitly suggesting that the pursuit of one will be fruitless.

Here I will ask the following questions: Are these two tasks as different as Arpaly claims they are? Does she give sufficient reason to conclude that a third person account of rationality cannot be translated into useful first-person guidance? Does she

---

[80] Arpaly, Nomy. 2006. (86-108) She argues this by analogy to belief. She claims that one who concludes not-Q from (P and P→Q) will not only be failing to follow a descriptive pattern of reasoning, but also will be doing something she ought not do, that is, making a mistake. Further, Arpaly argues, it will be a mistake whether the agent had the capacity to do otherwise or not.
[81] Arpaly, Nomy. 2003. (33)

offer a convincing argument that manual-style theories are altogether impossible? And, finally, does she offer any argument for the importance of prioritizing the quest for an account of rationality over the quest for guidance?

3.2.1 Rationality manuals versus accounts of rationality

In the passage above, Arpaly implies that people confuse the task of developing a rationality manual and the task of articulating an account of rationality simply because the distinction between the two is hard to notice, and is often overlooked. However, I think that urge to yoke these two pursuits together is not simply a matter of overlooking a salient difference. There is a straightforward argument to be made for the conclusion that possessing an account of rationality thereby provides adequate guidance for being rational.

P1) Possessing account R allows judge J the capacity, given full information, to determine whether any agent A is acting rationally or irrationally, as well as what A would need to do differently to act more rationally or perfectly rational in the given circumstances.

P2) Judge J is an agent.

C) Therefore, possessing account R will allow Judge J the capacity, given full information, to determine whether J is acting rationally or irrationally, as well as what J would need to do differently to act more rationally or perfectly rational in the given circumstances.

This argument demonstrates that, for the individual who is both a judge and an agent, having a third-person account of rationality is in fact equivalent to having a first-person manual for rational behavior. But there are at least three possible challenges to this argument, some of which Arpaly raises herself. I will take them in turn.

The first I will call the "know-how" response. This objection can be motivated by an analogy. Suppose that a botanist has a complete account of sequoias. Given full

information about any specimen, she can use the account to determine whether the specimen is a sequoia. Which is to say, for any specimen, she can know that it is or is not a sequoia. But, of course, she herself does not know how to become a sequoia. Nothing in the account of sequoias can tell her how to grow the appropriate limbs, how to turn her skin into bark, etc. And therefore, the objector may say, though the account is a perfectly sufficient third-person account of sequoias, it is nevertheless insufficient as a sequoia-realization manual. The bite of this objection comes in recognizing that, in the conclusion of the argument above, although Judge J may be possessed of clear judgment about exactly what changes she would need to make to become rational, the theory that allows her to make this judgment does not thereby empower her to make the necessary changes.

But this objection relies on unrealistic standards for instruction manuals. We do not expect instruction manuals to make us omnipotent. We expect them to tell us exactly what is required to realize a certain goal, and how we can use our actual physical capacities (perhaps with some practice) to bring this goal about. But it is not the fault of an instruction manual that it cannot give us new causal powers. We would not say that a blueprint for a house is instructionally insufficient because we do not have the wood or manpower to complete the project. Of course, in order to be effective in guiding behavior, every instruction manual must be combined with the goal of attaining what the instructions promise, and the capacity to do the things required to meet that goal. In the case of the botanist, it is not her instructional manual that is insufficient. Rather, she lacks the other two elements. She is not aiming to be a sequoia, and even if she were, it is anatomically impossible for her.

This leads to a second possible objection. It is true that we cannot hold our instruction manuals responsible when their efficiency is undermined by an agent's lack of relevant motivation or skills. However, we do realize that instructional manuals need to be intelligible to those whose behavior they are designed to guide. Consider the notation on a sheet of music that has been given to someone with both the goal of playing the melody it prescribes and also the skills to play the melody. Of course, the script might be unintelligible to her. She might lack the capacity to read music altogether, having learned to play completely by ear. Or perhaps the music is transcribed in some idiosyncratic and rare notation that only a handful of musicians worldwide understand. It seems clear that this is a shortcoming of the musical instructions qua instructions. Which is to say, while we do not take the sheet of music to be insufficient if it does not endow the musician with the desire and skills to play the song, we do take it to insufficient qua instruction manual if it is written in a manner unintelligible to the musician.

This highlights the fact that all instruction manuals must be audience-specific. The instructions therein must be communicated in a medium that the audience can grasp and translate into an understanding of what they need to do. And this leads to the second objection to the above argument, which I will call the "advice" response. As Arpaly suggests, in the case of rational guidance, it may be impossible to formulate a given instruction in a way that the agent can grasp and make use of. Consider the following passage:

> 'Never pay attention to anything I say' is incoherent advice, but it may still be true that a rational agent would never listen to anything I say…'Never act out of a conviction that is the result of self-deception'…is not very useful advice; neither is 'Never act out of an unconscious wish that has nothing to do with the task at hand.' Still, it may be true of the rational agent that she would never act out of such motives—that she would not simply try to avoid doing so, but in fact would

> never act out of them…'Act against your best judgment' is also an absurd piece of advice…Acting against one's best judgment is not something that one can, as a result of deliberation, resolve to do; one just changes one's best judgment as to what to do.[82]

The implication seems to be that, once we have a third-person account of rationality in hand, it may be incredibly difficult or impossible to convert this account into an intelligible set of instructions for how to be rational.

There are two reasons this response fails. First, while it may be that certain formulations of the account of rationality would be unintelligible as advice in the relevant sense, we cannot conclude from this that all formulations will be so unintelligible. Consider, for example, Liam. Liam is a compulsive liar who hopes to communicate this fact about himself to the world. Is it possible for Liam to communicate to us that everything he says is a lie? If he could make just one exception, he could of course say, "Everything I say, except for this statement, is a lie." But suppose for the sake of argument that Liam could never bring himself to make an exception. It might in this case be impossible for him to communicate to us that he never tells the truth. But of course, the fact itself can still be communicated. A non-liar, a sign hung around Liam's neck, the regular discrepancy between Liam's claims and our own observations—any of these could alert us of the fact which Liam himself cannot communicate.

Arpaly imagines the instruction manual must come in the form of general rules written directly from the third-person account. But why, we may wonder, would the advice need to be framed in this particular way? If, given a complete account of rationality, I know that the rational thing for you to do is spend the afternoon studying, even though you have decided in accordance with your best judgment to go to the park,

[82] Arpaly, Nomy. 2003. (34-35)

why must I communicate this to you by saying, incoherently, "Do the opposite of what you decide to do."? Why can I not advise instead, "Stay home. Study. Do not go to the park."?[83] Even if some ways of translating a third person account of rationality into instructions fail to yield an intelligible manual, this does not mean all such attempts are similarly doomed.

But there is an additional reason the advice response fails as an objection. Since, we are imagining, the person who is in possession of an account of rationality is herself the agent in need of guidance, communication becomes moot. If I know in advance that Liam is incapable of telling the truth, any blunders that ensue when he tries to communicate this point to me are irrelevant. Similarly, if I have a complete account of rationality, I will not need to translate the account into communicable advice in order to use it as a guide. It will show me directly what the rational thing is for me to do, and I will do that.

Finally, in response to this defense, it could be argued that, in interpreting the above passage as having to do with forms of communication and advice, we somehow miss the deeper point, which has not to do with communication between the individual who possesses the account of rationality and the individual who needs a manual for rational guidance. Rather, what is lurking here is a problem of <u>access</u>. There is, this response maintains, something about being in a scenario in which you need guidance that guarantees you cannot use a third person account of rationality as the source of that guidance. I will call this the "access" response. Arpaly writes,

[83] In fact, I believe there is a good reason that the latter advice might not be an effective guide to your behavior. But this has nothing to do with whether it is intelligible, but rather whether you trust me as an authority on what is rational, a point I will return to at great length later.

> It might even be possible that just as the right account of delusion must imply that one cannot tell if one is deluded or not, the right account of rationality must imply that, at least in some situations, one is a very bad judge of one's own rationality.[84]

In response to this suggestion, we might wonder first whether it is true that the right account of delusion must imply that one cannot tell if one is deluded. But granting that it is, then in order to preserve the analogy the first premise in the original argument must be revised to say:

> P1) Possessing account R allows judge J the capacity, given full information, to tell whether <u>any agent A except herself</u>, is acting rationally or irrationally, and to what degree, and what A would need to do differently to act more rationally or perfectly rationally.

Let us suppose for the sake of argument that the right account of delusion is the one that equates being deluded with having false beliefs. From a third-person perspective, given full information, this would allow us to say of any given agent whether or not that agent is deluded. For instance, suppose one year ago we knew 1) Bernie Madoff believed his scheme was never going to be exposed, and 2) Bernie Madoff's scheme was going to be exposed, we would rightly conclude that Madoff was deluded. One of the things we <u>go looking for</u> when trying to discern whether a given believer is deluded is a discrepancy between his beliefs and reality.[85] Given that we can use this account of delusion to diagnose others, can we not thereby use this account of delusion to diagnose ourselves? There is reason to think we cannot.

First, while it is often easy for a judge to detect the discrepancy between another's beliefs and reality, we might think it is rarely easy, perhaps impossible, for a judge to spot discrepancies between her own beliefs and reality. Of course, in the previous

---

[84] Arpaly, Nomy. 2003. (34)

[85] A complete account of delusion involves more than just—false beliefs, for instance, a persistent unwillingness to revise beliefs in light of contrary evidence.

chapter I discussed how checking for error is a regular component of human mental activity. If I am going to the store and I believe I am out of milk, I might check the refrigerator to be sure my belief is correct. In this case, I can use direct perception of a state of affairs to check the accuracy of my beliefs about that state of affairs, and, often revise my beliefs in light of this perceptual evidence. Surely this is a case in which I have spotted a discrepancy between my beliefs and reality.

But I gather that the objection runs deeper than this. The objection is grounded, I suppose, in the realization that all the double-checking we can do must be done from our first-person perspective. And some types of evidence of error that are very difficult or impossible to see from this perspective are nevertheless easily seen from an external perspective. Hence, we can use our account of delusion to judge others, but not ourselves.

But realize here that the reason we cannot translate our account of delusion into a delusion-avoidance manual is not because of some insufficiency in the account of delusion, but rather with our lack of full information. The response gets its force from our intuition that it is much more difficult, impossible even, to possess the information about ourselves that we need to apply the account of delusion to ourselves than it is to possess the information about others that we need to apply the account of delusion to them. If we were capable of attaining full information about ourselves, one of the things we would learn is whether our beliefs are false, and hence, whether we are delusional.

Later in her discussion, Arpaly writes, "I take it as an advantage of my view [of rationality] that it correctly predicts a world in which people often have to ask others to

judge their own rationality, 'Am I crazy?' they ask."[86] Let Ann be an agent, and Judy be an observer, and let's suppose Ann asks Judy, "Am I deluded here?" It seems to me that in order to ask this question, Ann presumes that Judy more or less shares her third-person account of delusion, but Ann suspects that Judy has information that she herself does not have, information that can be more easily, or perhaps exclusively, gleaned from an external perspective. If this is the case, then what stands between the third-person account of delusion and a perfectly adequate first-person delusion-detecting instruction manual is nothing. Rather, Ann's lack of full information is creating the gap.

All of this discussion leads to the following insights. A complete third-person account of rationality is not alone sufficient to guide us in rational behavior. For, in addition to such an account, what is needed is the following three things: the active goal of being rational, the skill to do the things that lead to the fulfillment of this goal, and access to the full information regarding whether one is doing the things that lead to the fulfillment of this goal. But notice, none of this suggests that we need a <u>different</u> account of rationality to serve as an instructional manual. Rather, the third-person account must be accompanied by certain motivational and skill-related conditions in order to serve as an instruction manual.

Hence, Arpaly's original suggestion that there are two independent lines of inquiry regarding rationality—one that tries to provide an instruction manual and one that tries to provide a third-person account—is undermined. Of course, the person who seeks to be rational may have problems that the person who seeks merely to judge whether others are rational does not have. She may face limitations on her own skill set and motivational capacity. She may also face persistent worries about the reliability of the

[86] Arpaly, Nomy. 2003. (61)

information she uses to judge whether she is being rational at any given time. But these difficulties do not arise because her account of rationality is insufficient as a guide. They arise because the world conspires against her using the guide to change herself.

Returning to the original dialectic, we must wonder how these insights bear on the defense of guidance-first approaches to normativity. In demonstrating that Arpaly's distinction between accounts of rationality and instruction manuals for rational behavior is indefensible, perhaps I have proved too much. For, it appears the best we can do if we want an instruction manual for being rational is to locate a complete third-person account of rationality, and then work hard to make sure we have the skills and knowledge needed to apply the account to ourselves.

In other words, the conclusion from the above discussion seems to be that, to the extent we are interested not only in distinguishing rational from irrational behavior but also in being rational ourselves, we thereby simply give ourselves extra non-ethical work—namely, the epistemological work of making sure we have full relevant information and the practical work of developing whatever skills it takes to be rational. It is reasonable to think that this extra work is not within the purview of ethics. Which is a way of arguing that only the first, non-guidance task—the task of providing a third-person account of rationality—should be the concern of ethicists and those worried about practical rationality.

Has our defense of guidance-first ethical inquiry failed? In realizing that the two tasks that Arpaly takes to be distinct are in fact not distinct, have we thereby inadvertently made her dismissal of guidance-first approaches all the more inevitable?

In the next section, I will argue that we have not. For, if we take the lessons of prospectivism into consideration, if we remind ourselves that deliberators have no indubitable ends, not even the end of being rational, we begin to understand why the ethicist must do more than provide mere third-person accounts of the concepts—morality, rationality, bravery, patriotism, kindness, etc.—that guide our behavior by serving as ideals. Even further, we begin to understand why any accounts of ethical concepts that are formulated without concern for an agent's guidance cease to be binding concepts for that agent at all.

3.2.2 Begging the question

Conceptual analysis is the bread and butter of philosophy. We philosophers analyze concepts better than anyone else, and we cannot seem to help ourselves. Our friends ask us, "Do you like the way the Obama administration is handling the financial crisis?" and our first response is, "Well, it depends on what you mean by 'like'…" It is no surprise then, that we cannot resist the urge to analyze some of the most prominent action-guiding concepts in our conceptual arsenals—good, moral, rational, right, valuable, important, etc. We want for each of these a list of criteria that will help us distinguish for any given agent, including ourselves, whether that agent's actions fall into the relevant category.

But, when we are deliberating sincerely about what to do, often the question we seek to answer is not primarily a question of conceptual analysis, nor a question of applying concepts to determine which actions would fall into which categories, though there is certainly some of this going on. Rather, the primary question is which categories our actions ought to fall into.

In the case of descriptive concepts, like 'walking' or 'jumping', our concepts themselves do not purport to answer this question, any more than our concept of 'unicorn' purports to answer the descriptive question of how many unicorns exist. Rather, it is our desires, our norms, our value judgments, and the commands of others, which, employing these concepts, purport to answer the ought question. When we accept the authority of these sources of prospective content, we accept as binding the normative answers they give regarding what is to be done. In these cases it is easy to spot the difference between the following questions:

1) Analysis and application question: What is required for me to be walking?

2) Authority question: Is walking now required of me?

Things get trickier, however, when the concepts we are analyzing are normatively loaded concepts like moral, rational, right, brave, etc. These concepts, helpful for allowing an individual to keep track of what is to be done and what is not to be done, wear their claim to authority on their sleeves. Korsgaard writes,

> Every moral theory defines its concepts in a way that allows us to say something negative about people who [challenge or ignore claims involving those concepts]—say, that they are amoral or bad. But an agent who doubts whether he must really do what morality says also doubts whether it's so bad to be morally bad...[87]

The problem with the concept of morality is that it blurs the distinction between the analysis and application question and the authority question. Because an analysis of 'moral' involves, as a defining criterion 'required', it blurs the difference between the questions:

1) Analysis and application question: What is required for me to be moral?

2) Authority question: Is being moral now required of me?

[87] Korsgaard, Christine. 1996. (16)

The guidance-first approach to normative inquiry can be understood as an antidote to this kind of confusion. It insists that these two questions are distinct and need to be treated as distinct, and that authority questions are the primary questions at stake in general ethical inquiry.

The above discussion of Arpaly's distinction between accounts and manuals allows us to see that those things that are required to turn an account of rationality into an effective 'rational behavior instructional manual' are not the type of things that ethicists can, or should try to, provide. But it does not thereby follow that ethicists cannot or should not try to answer authority questions. For these questions are never of the form, "What is required to achieve goal g?" Rather, the general form of the question is: Given that I cannot take all of my antecedent prospective content as authoritative, which prospective content, and which sources of prospective content, should I take to be authoritative?

Korsgaard writes that the "tendency to conflate the normative question with other questions often results in the normative question being blocked or ignored."[88] Indeed, it seems Arpaly has done just this. Particularly, working with what she believes to be an intuitive analysis of rational, she then argues that any action that does not fall into the category is not merely not-rational, but it is in addition, mistaken.[89] By successfully establishing that the guidance needs of the sincere deliberator are irrelevant to the question "what is rational?" when that question is treated as a descriptive analysis and application question, she takes herself to have established that the needs of the sincere

[88] Korsgaard, Christine. 1996. (43)
[89] See fn. 80.

deliberator are also irrelevant to the related authority question. But the latter does not follow from the former.

3.3 Misunderstood Deliberators

3.3.1 Detecting rationality

> When asked directly why she believes deliberation is important, Arpaly writes,
>
> > Our higher faculties, it seems, are a wonderful asset when it comes to dealing with new information and unfamiliar data…the ability to deliberate and reflect helps make us, as humans, be rational much more often, about many more things, in many more contexts."[90]

In particular, she believes, deliberation helps us be more rational by endowing us with the following advantageous capacities: the ability to reconsider our intuitions in unfamiliar scenarios, where our intuitions are often unreliable; the ability to double-check our gut instincts when we have become aware that our gut instincts on a particular matter often lead us astray; and the ability to focus our concentration on competing considerations regarding any weighty matter.[91]

> Arpaly acknowledges a lurking tension here, admitting,
>
> > Some of the [advantages of deliberation] on my list might make one wonder if I am not taking back some of what I said before. How do these work if deliberation does not have the kind of authority that some action theorists invest in it…?"[92]

She responds by writing,

> > But this is like saying that it makes no sense to install Norton Utilities on my computer's hard drive simply because Norton Utilities is a fallible computer program that occasionally erases a good file. All in all, deliberation increases our chances of being rational. Occasionally, however, it only stands in the way of reasonableness.[93]

---

[90] Arpaly, Nomy. 2003. (64-65) Emphasis hers.
[91] Ibid. (63-64)
[92] Ibid. 2003. (65)
[93] Ibid. 2003. (65)

How is this analogy supposed to defuse the charge that deliberation does have genuine authority? The software picture is this: There is a fact about which files are bad for one's computer. Agents run Norton Utilities software because, wanting to avoid these files, they aim to discover which of the files on their computer are bad. When the software locates these files, it destroys them. Users trust the software's authority because they believe that, more often than not, it destroys the bad files. The analogous deliberative picture is this: There is a fact about what it is rational to do in any given circumstance. Agents deliberate because, wanting—consciously or unconsciously—to act rationally, they aim to discover which of all the actions available to them is the rational action. When they locate the rational option, they act in accordance with that option. Agents trust deliberation's authority because they believe that more often than not, it delivers the rational action.

Against this backdrop, Arpaly's point is simple and quite compelling: Our deliberative rationality-detecting systems, just like our utilities software, are fallible. In addition, sometimes our instincts will lead us to do the rational thing even if we do not deliberate, just as sometimes our computers will remain in good working function if we do not run Norton software. We are warranted in trusting deliberation's conclusions, Arpaly believes, because far more often than not, they deliver accurate answers to the analysis and application question, "What is rational?"

This may be a satisfactory answer to action theorists who believe the constitutive aim of deliberation is deciding which course of action is rational. These theorists may thereby be persuaded that the authority of deliberation is dependent on its rationality-

detecting capacity in the way that the authority of Norton software is dependent on its virus-detecting capacity. But, as I have argued at length, the constitutive aim of deliberation is not rationality.

There is another reason that philosophers, including myself, might "invest deliberation with authority" that Arpaly overlooks. Deliberation, I maintain, does not have authority because it yields some privileged access to facts about what is the rational, good, or right thing to do, facts that the deliberator would otherwise lack. Rather, deliberation has authority because, when someone is genuinely deliberating, her deliberative processes are the only things she can take as authoritative, if she is to take anything as authoritative at all. Another analogy might help make this distinction clear.

When survivors of a shipwreck climb aboard the only remnants of the wreckage—a tattered, leaking lifeboat—they do not trust it to carry them to safety because they believe it possesses the qualities that, generally, allow it to serve that function. Rather, they trust it because it is their only option. It is likely, in fact, that they have grounds for doubting the capacity of the lifeboat to deliver them to safety, and yet trust it with the task anyway. Similarly, deliberators may doubt, in light of Arpaly's discussion (and, it is worth mentioning, a heap of empirical evidence from moral psychology that supports the conclusion that conscious reflective deliberation often makes us worse rather than better off[94]) that their deliberations will steer them toward well being, or any other generally desirable state. But trust their deliberative processes they must.

Consider now the judges who are equipped with a full grasp of all relevant concepts and full information about the world, and who are watching these deliberators

[94] See, for example, Lehrer, Jonah. 2009. And, Gladwell, Malcolm. 2005.

from the third-person perspective. Whatever these deliberators do, the judges will be able to discern exactly which categories they and their actions fall into. The judges will, because of their privileged perspective, accurately judge whether the deliberators have done something kind, beautiful, rational, moral, good, funny, charming, useful for society, etc. And they will also know whether the deliberators have been deceived in believing that they have done any of these things. If the deliberators are so deceived, they are surely mistaken in that sense.

But it is not a mistake for these deliberators to trust their deliberative processes. They have no choice in the matter. Short of waiting for some outside force to make up their minds for them, deliberation is their only hope for determining what is to be done. The question thus becomes, exactly what are deliberators trusting their deliberative process to do during deliberation?

3.3.2 Awareness of our mental states

One seemingly unique talent humans possess is the ability to be aware of, and attend to, our own mental states as such. This talent is the foundation of reflective self-awareness, and it is taken to be crucial to our capacities for reasoning and deliberation. Beyond this, however, some philosophers have suggested that the awareness of our mental states creates a problem for us that no other animal has, a problem that reflective deliberation is then employed to solve. The following two passages, from Korsgaard and Nagel, respectively, attempt to articulate exactly what problem is caused by our ability to attend to our own mental states.

> [T]he human mind is self-conscious in the sense that it is essentially reflective…A lower animal's attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious of them. That is, they are not the objects of its attention. But we human animals turn our attention on to our perceptions and desires themselves, on to our own mental activities, and we are conscious of them. That is why we can think about them. And this sets us a problem no other animal has. It is the problem of the normative…I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now that impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a reason to act? The reflective mind cannot settle for perception and desire, not just as such. It needs a reason.[95]

> The capacity for self-consciousness changes the nature of the being who is making the choices—whether they are decisions to act or decisions to believe—by introducing irrevocably the distinction between appearance and reality, between how things seem from our personal point of view and how they really are—and facing us with the need and perhaps providing us with the capacity to arrive at an answer that seems right not just from our individual point of view but from the reflective standpoint that takes this view as its object.[96]

To return to the question at hand, these theorists believe that, during deliberation, deliberators trust their deliberative processes to resolve the problem (or, alternatively, meet the need) that arises as a result of our reflective capacity. The question I want to ask here is whether the problems confronting the individuals described by Korsgaard and Nagel are the same as the problems that confront the prospective deliberator. I will conclude that deliberators rarely, if ever, face the problems described by Korsgaard and Nagel, but that it is easy to mistakenly think they face such problems because the above problems are similar to the problems that are in fact regularly faced by prospective deliberators. Further, to the extent that any individuals actually do face the above problems, there is nothing we can do, as philosophers, to help them.

3.3.3 Aiming for certainty

[95] Korsgaard, Christine. 1996. (93) Emphasis hers.
[96] Nagel, Thomas. 1996. (202)

If prospectivism is true, we can immediately spot a flaw in Nagel's understanding of the reflective agent's need. For, it is not self-consciousness that introduces the distinction between appearance and reality. The mere recognition that we have a belief is not itself enough to challenge the belief's accuracy. Rather, it is evidence of error that first makes us aware of the possible discrepancy between what we believe and how things actually are. It is, for example, the moment we realize how many houses there are in the world, that we begin to question whether any man, no matter how many flying reindeer he has at his disposal, could possibly visit all the houses in one night. And then we might wonder: Does Santa have the power to stop time? Does he deliver presents to other countries on different days? Does he delegate some of his work to our parents? Etc. It is not first the capacity to be aware of beliefs as such that introduces a problem to be solved. Rather, awareness of incompatibility among beliefs introduces uncertainty that cannot be resolved until we modify our belief set.

In the practical realm, not only are our beliefs called into question, but also our goals, our desires, and the norms that we use—consciously or unconsciously—to govern prospective content. For the sake of discussion, I will say that when a belief is in doubt its accuracy is in question; and when a norm is in doubt its authority is in question. Using this terminology, we can say that during regular, non-reflective activity, we take our beliefs to be accurate guides to how the world is, our prospections to be accurate guides to how the world will be, and the norms and desires that shape the content of those plans to be authoritative guides to how the future should be.

Allow me to make an important note about this terminology. To insist that we take our norms and desires to be accurate as opposed to authoritative guides to how the

world should be is to smuggle in a moral realism that we do not need and that causes innumerable problems. I am not here rejecting moral realism outright, but I want to ensure that my vocabulary does not implicitly endorse the position. Similarly, as the practical analog of the epistemic notion of evidence of error that arises when there is some problem with one's belief set, I will use the notion of cause of uncertainty to describe a situation in which a problem arises with one's prospective set. Cause of uncertainty threatens the authority of the various sources of, and constraints upon, prospective content.

Prospectivism maintains that it is not in the first place general doubt that causes the problem about which we must deliberate, but rather a particular incompatibility among beliefs, plans, and norms that causes problematic uncertainty about which we must deliberate. And while we deliberate, it is our reflective awareness of the beliefs and norms at issue, our ability to bring these mental states into focus, and in particular, our ability to articulate and grasp exactly which norms and beliefs are called into question by a particular development that contribute to our capacity to self-correct, i.e., to process and neutralize the cause of uncertainty. Reflective awareness in this case is not the source of a problem; it is part of the solution.

Still, we might think Nagel's important point stands independently of the question of how we initially become aware of the possibility of error. What really matters, we might suspect, is that the threat of error, once introduced, is introduced irrevocably. Having learned your belief in Santa Claus was inaccurate, you have thereby been made aware that beliefs are, in general, the types of things that are susceptible to inaccuracy. Similarly, if you have ever consciously surrendered a norm because you came to

recognize it lacks authority, you have thereby been made aware that any of your norms may lack authority.

The awareness of the possibility of error is the birthplace of skepticism. Those who are profoundly bothered by the possibility of error set out to put it to rest, Descartes providing perhaps the most famous example. It is important to highlight that such thinkers are not aiming to overcome particular evidence of error. Rather, they want to eliminate the possibility of error altogether. In the practical realm, achieving this task would involve ruling out any possibility that a given norm lacks authority. Is this what deliberators are up to when they stop and ask themselves what they have reason to do? Are they checking to see whether the norms guiding their plans are backed by an indubitable source of normative authority? For such deliberators, does the need for guidance always involve the need for immunity from uncertainty?

Before answering this question, let us return for a moment to the prospective deliberator as a point of contrast. Prospectivist deliberation begins not with an agent's awareness of the <u>possibility</u> that her prospective content is flawed, but rather with her awareness that her prospective content <u>is in fact</u> flawed. The hiker from §1.3.5 cannot plan both to keep her feet dry and to reach the campground by nightfall. The pregnant teenager cannot plan both to obey the teachings of her church and to finish high school on schedule. The economically struggling family cannot both pay their mortgage and pay for health insurance. In these situations an agent becomes aware that something <u>must</u> give. In the case of belief, an agent must somehow make the beliefs compatible by revising or surrendering one or all of her competing beliefs. In the case of prospective content, an agent must somehow make her goals compatible by revising or surrendering

one or all of the goals (and perhaps the norms constraining those goals). This is to say, prospective deliberation is spurred not by an agent's awareness of the threat of error, but rather because the agent has direct cause of uncertainty.

It is possible, of course, to turn a prospective deliberator into a practical skeptic. All you need to do is add to an agent's prospective set the activated goal of being certain that all her beliefs are accurate and all her norms are binding. But then we must ask what would satisfy this demand for certainty. We might think that such an aim requires us to try, as Descartes did, to locate some indubitable foundation for all our beliefs and norms, to find a belief that must be accurate and a norm that must be authoritative. But, realizing the futility of this pursuit—and I maintain that such a pursuit is indeed futile—the agent might instead do the most thorough scan she can for any cause for uncertainty, and finding none, be content that she is as certain as anyone in her situation can be.

In light of these two possibilities, I argue the following: If we are meant to understand Korsgaard's or Nagel's deliberator as engaged in the same pursuit as the first agent above, as looking to overcome practical skepticism once and for all, we philosophers should not concern ourselves with her problem, for two reasons. First, it is unlikely that there are any actual agents who possess the goal of absolute eradication of the possibility of uncertainty. This is an obviously false description of actual humans. Even Descartes trusted his belief that his pen existed enough to use the pen to write about whether he could trust that his pen existed. Second, even if it is psychologically possible to adopt such an aim, and even if a handful of dedicated thinkers have adopted the aim, it is impossible to satisfy. There is always a fanciful what-if waiting in the wings—a demon, the matrix, delusion-inducing drugs—raising the possibility that our

understanding of the world and ourselves is flawed. Since the task of eliminating this possibility is impossible, we should rightly conclude that nothing we philosophers can come up with will satisfy this aim.

3.3.4 Checking for uncertainty versus deliberating

Thus, being charitable, we should assume that, insofar as a reflective agent is after some sort of reassurance, it is the latter type, the certain-as-I-can-be type. And here we arrive at an important distinction. The activity of practical reflection and practical deliberation are often treated synonymously. But notice, the agent who, motivated by a desire for certainty, goes looking for cause of uncertainty is not yet deliberating. Rather, she is checking whether she has anything to deliberate about, checking to make sure that she has not overlooked accessible cause of uncertainty.

The habit of checking for accessible cause of uncertainty (henceforth, "checking for uncertainty") is—unlike the aim of attaining absolute certainty—very common. Causes for uncertainty are sometimes easy to miss, even when they are staring us in the face. And since the consequences of missing important cause of uncertainty can be dire, we develop habits of actively searching for them, and expect others to do the same. We expect our students to recheck their grammar before they submit papers. We expect our pilots to check that the instruments are properly functioning before take-off. We expect the doctor to check he is amputating the correct limb before he begins his incision. And we hold these agents responsible when their failure to engage in such checking results in catastrophe (or, in the less extreme case, results in an unreadable paper).

Checking for uncertainty is thus an important and perhaps uniquely human form of reflective activity. In the practical case, checking for uncertainty can sometimes itself

involve imagining alternatives to one's current plan in an effort to make sure there is not some more attractive option. It can also involve consciously attending to one's normative beliefs to make sure that the current plan accords with those beliefs. And it can involve imagining the current plan in greater detail to check for lurking incompatibilities. But, importantly, unless any of these searches turns up an actual instance of conflict, an agent has nothing to deliberate about. For, deliberation is what is required once the uncertainty is discovered.

When the student realizes her paper is riddled with incomplete sentences, when the pilot realizes that his gauges are not functioning properly, when the doctor realizes that the leg he is about to amputate does not match the description on the chart—these are the scenarios that call for deliberation. How should I revise my sentences? How serious is the malfunctioning of these instruments, and what are my options now that I am aware they are not properly functioning? Whose chart am I looking at, and why was it handed to me? Thus, our reflective nature comes in two phases: the checking phase and the deliberating phase. If the checking phase turns up no uncertainty, there is nothing for an agent to deliberate about.

3.3.5 The power of reflection

This distinction can help us shed light on what is misleading about Korsgaard's understanding of the deliberator. Korsgaard suggests that when an agent attends to her mental states, the attention itself has the effect of neutralizing the causal impact of those mental states. Korsgaard characterizes this as the transformation of incentives—objects represented in such a way as to motivate us directly—into inclinations—proposals

offered up to reason as possible options.[97] We can illustrate this idea with an analogy. The difference between an incentive and an inclination can be compared to the difference between a wind that moves a ship, and the same wind conceived by a skilled sailor as a possible source of movement of the ship. The wind, possessed of its own direction and force, provides a suggestion or proposal to the sailor regarding where he might sail, but the sailor is in a position to decide whether to engage the wind or not.

The story is similar with an agent who, say, recognizes in herself an inclination to run away in the face of some danger. Korsgaard believes that the recognition and reflection upon this inclination itself suspends its direct casual efficacy, and transforms it into a proposal that the agent can take or leave. She writes, "In the sense I am using the term here, an inclination just is the operation of a natural incentive upon you, as viewed through the reifying eye of self-consciousness."[98] Faced with the inclination to run away, the agent must decide whether to run away.

As an initial response to this conception of deliberation, it is worth noting that in fact, often when we attend to our desires, it makes them more not less dominating. Thinking about the fact that one has the urge to laugh in church often makes the impulse to laugh stronger, not weaker. Reflecting on one's hunger or on a particularly appealing food item usually increases the demand for food. This evidence seems contrary to the claim that attending to the incentives motivating us at any one point suspends their causal efficacy.

In contrast, prospectivism does not take reflection upon an incentive to suspend that incentive's direct causal efficacy any more than a sailor's reflection on a wind

[97] Korsgaard, Christine. 2009. (121-122)
[98] Ibid. (121-122)

suspends its causal efficacy. Rather a problem with the prospective content related to the incentive is what disarms the incentive. In other words, incentives interrupted by a particular conflict are brought to our attention not merely as inclinations that we can take or leave, but rather as problematic inclinations that, in their current state, cannot move us.

However, if one overlooks the difference between checking for uncertainty and deliberating, and groups them both under the heading "reflecting," it might be easy to mistakenly think that reflection itself has the power to suspend the efficacy of desire. Normal human adults often, as part of a useful habit, stop themselves to check for uncertainty. And this process, which involves attending to the content of our beliefs and plans, even occasionally attending to the content as such, involves a momentary suspension of our confidence in our beliefs, plans, and norms, and hence an openness to uncertainty, i.e., to evidence that the norms and desires shaping our prospective content are not authoritative. Korsgaard mistakenly attributes to agents suspended in such a search the need for reasons. But in fact, the reflection is centered on locating cause of uncertainty. We do not need uncertainty; we do not even in most instances want it. Quite the opposite, we are usually hoping we will not find any. For any such uncertainty is inconvenient. It requires us to deliberate. It is only once we have found uncertainty that the need for reasons arises.

3.3.6 Reflective hang-ups

In previous chapters, I have tended to describe those scenarios that lead to deliberation as scenarios in which uncertainty finds us. If readers have sensed that this is an inadequate account of deliberation, it might well be because they think, rightly, that we do and ought to deliberate more often than just in those scenarios in which we

stumble upon incompatibilities in our prospective set. Understanding that many of the causes of uncertainty that spur deliberation are discovered not by accident but because, as avid critical thinkers, we go looking for such incompatibilities, allows us to preserve prospectivism as a general theory of deliberation while accounting fully for the reflective activity of even the most thoughtful of agents.

To summarize, Korsgaard claims that agents, while reflecting, are looking for reasons to do things. Nagel believes that they are hoping to quiet the threat of possible error. But prospectivism maintains that when agents are actually deliberating, it is because they are dealing with some actual cause of uncertainty. Korsgaard writes, "[The reflective mind] needs a reason. Otherwise, at least as long as it reflects, it cannot commit itself or go forward."[99] Prospectivism counters with a similar, but importantly different claim about the needs of the reflective self: The reflective mind, when confronted with uncertainty, needs to process and neutralize this uncertainty. Otherwise, it cannot commit itself or go forward.

These observations raise a number of questions. How can an agent neutralize uncertainty? Are some methods and instances of neutralizing uncertainty better than others, and if so, by what standards are they better? And, meta-philosophically, is there a role for philosophers to play in answering these questions?

[99] Korsgaard, Christine. 1996. (93) Her emphasis.

Chapter Four

The Vindication of Values

4.1 Framing the Authority Question

If you are about to take a bite of soup and a fly lands on your spoon, your prospective content becomes unsettled. You have an aversion to consuming any food that has been touched by a fly, but you have a strong desire to eat. The question is whether you should override the aversion and allow yourself to take a bite, or whether you should override your desire to eat instead. This tension can be resolved by deliberating to discover a revised plan to fetch a new bowl of soup. In doing this, you respect the authority of your aversion to fly-tainted food, while still allowing your hunger to shape a new plan, forsaking only the immediate desire for a particular bite of soup.

Of course, this is a fairly trivial example. A much more weighty decision might arise in the life of a pregnant teenager who believes abortion is wrong but who is considering aborting her fetus before telling her community that she is pregnant. Her general value judgment that abortion is wrong serves to prevent the teenager from accepting as her future any plan in which she has an abortion. But carrying a child to term is incompatible with all her current plans, which are forged from various desires and value judgments (getting good grades in school, playing on the volleyball team, maintaining her social status and friends, protecting her family's reputation, etc.).

Here, the anti-abortion aversion is much different in nature than the anti-fly aversion. It is imbedded in a large and stable identity, a network of norms that purport to

direct an agent toward a well-lived life, and further, direct an agent how to feel about herself upon fulfilling or failing to fulfill these norms. Which is to say, parts of the agent's identity—the honest daughter, the good Christian, the compassionate, life-affirming citizen—are in conflict with other parts of her identity—the good team-member, the talented student, the fun and carefree friend, the girl who lets nothing get in the way of her dreams.

This is why our decisions are often not just about what to do, and what to value, but also about who to be. When we knowingly override our aversions to pain or discomfort—when we make ourselves go to the dentist because it is the only way to fix our teeth, or we make ourselves eat the fly-tainted soup because it is the only food we have in the house—it is simply a matter of bringing ourselves to endure, temporarily, an uncomfortable experience. But when we override the aversions and desires that stem from our value systems, we do not just agree to endure temporary discomfort; rather, we agree to revise our identities.

It is important to note that although the weightier, value-laden decisions we make differ in significant ways from the more pedestrian fly-in-the-soup decisions, they are not different in kind. For, at issue in any open-ended practical deliberation is the question of import. Is it important to avoid eating the fly? Is it important to satisfy my desire to eat? Is it important to avoid having an abortion? Is it important to be a good Christian? How important is it and in what sense is it important?

In order for diagnostic reason to answer these questions, it must have access to a working theory of value—a theory that tells a deliberator what is important and why. Included in this is a working theory regarding which sources can be trusted as authorities

on the matter of import. Our desires, our learned norms, our value judgments, pain and pleasure, the demands made by other people, the yelps of a suffering dog—all of these things direct our attention to objects, beings, and situations. Even just the fact that other people are attending to something itself suggests that we should pay attention to it. In this simple way we are continuously bombarded by demands for attention and specific, related action. (These demands, when unreflectively guiding action, are what Korsgaard calls "incentives.")

One component of a working theory of value is a theory of which of these demands can be justifiably discounted or ignored. The recovering smoker might discount his desire for a cigarette. The sexist might discount the requests of women. The factory farm owner might discount the cries of suffering animals. Our theories of value often take the form of generalized policies regarding which of the various demands for attention and action need to be heeded and why. These judgments are what I have been calling, for shorthand, judgments of import.

Thus, the authority question is properly framed not just as uncertainty with regard to the authority of one given demand for attention and action, but rather as uncertainty about which components of a set of conflicting or vague demands should be granted authority as is, which should be granted authority in some revised form, and which should be ignored temporarily or altogether. When we inquire as to the philosophical prospects of answering the authority question, then, we are inquiring whether philosophy has anything to say about this collection of questions. Further, the authority question is always going to be found in the context of a conscious attempt to reevaluate in the face of uncertainty about the authority of one or more demands.

Another important point is that deliberative uncertainty is not caused by momentarily forgetting one's value theory. If that were the case, all an agent would need to do is recall to mind her theory of value, apply it, and move forward. Deliberative uncertainty is caused by a shortcoming with the value theory itself. As a result, in the face of such uncertainty, we cannot simply rely on the rules, norms, maxims, desires, identities, etc. that we previously took as authorities on the matter of what is important, for these entities are playing a direct role in causing our deliberative problem.

This observation reinforces the point made in the previous chapter that even the clearest analysis of a given action-guiding concept (e.g., rationality) will not be able to instruct us in our deliberative revisions so long as its authority is what is in question. Similarly, we cannot turn to any fixed set of rules, or any rigid identity that we have antecedently taken to be an authority on what is important as a means of solving our problem. Even in the case in which we decide to trust a given person (like a parent, friend, or pastor) as an authority on a given problem we are facing, doing so requires us to extend our understanding of that person's authority, to judge explicitly that they have authority on the specific matters in question. Deliberation, I insist, calls upon us to change our value theory via extension, specification, or revision.

4.2 The Authority of Experience

Prospectivism allows us to make sense of the idea that developments in one's life and one's environment can cause problems for one's normative theory. Here, I will offer and defend an account of the conditions under which revisions to one's normative theory can be vindicated, an account I will call the evidential account of normative authority.

Succinctly, the position is this: An agent's normative theory—the collection of her judgments of import—is vindicated if guiding her actions in light of the theory proves to be non-problematic. Further, evidence of authority is accrued over time in relation to whether the normative theory proves non-problematic not only in a given instance but rather in a variety of environments and motivational circumstances.

4.2.1 Cause of uncertainty versus grounds for uncertainty

One of the first skills a philosopher must learn is the ability to appreciate the distinction between belief and justified belief, i.e., between actually believing and having grounds for belief. Those who seek to defend rational desire are committed to a similar idea, the distinction between actually desiring and having grounds for desire. These distinctions, in turn, anchor our judgments that certain actual beliefs and desires are mistaken.

We might expect a parallel distinction to be made between actual uncertainty and grounds for uncertainty. And this leads to an obvious objection to the evidential theory of normative authority. If the evidential theory grounds the vindication of value adjustments in their capacity to resolve actual uncertainty, but agents are capable of being unwarrantedly certain or uncertain, then it seems the wrong norms will be vindicated on this picture.

For example, the norm of enslaving my friends in my basement and forcing them to write articles for me would be vindicated so long as it causes me no uncertainty. This seems, initially, as unattractive as the proposition that my desire to enslave my friends justifies my plan to do so. It also might lead to the inverse counterintuitive consequence. Suppose I am such that the goal of basic hygiene causes me a great deal of normative

uncertainty: I just cannot convince myself that bathing is worth the bother. In the face of this uncertainty, I will be forced to revise my norms. But, we may believe, since the uncertainty was ungrounded in the first place, so too the vindication of any revisions made to accommodate the uncertainty is similarly ungrounded.

These are the types of scenarios that lead Arpaly and others to reject the idea that whether we can convince a given moral agent to accept a given value has no bearing on whether that value is authoritative. Such a rejection is comforting for two reasons. It forms a nice analog with our typical rejection of the idea that whether we can convince a given believer to accept a given belief has no bearing on whether that belief is accurate. It also allows us to dismiss without much hesitance the authority of the norms of others, when those norms seem suspect or unjustifiable. If, for example, you are walking down the street and a stranger demands that you must publicly accept Jesus Christ as your lord and savior, one option available is to reject the stranger's demand as unjustified. He is certain about it, but this is only because he has failed to appreciate the abundance of actual evidence that would undermine his certainty. In general, the proposal that freedom from uncertainty is equivalent to having justified values and actual uncertainty is equivalent to having unjustified values is unattractive. If the evidential theory of normative authority commits us to this position, we ought to reject it.

In response to these worries, I nevertheless maintain that there is nothing more to a given value's vindication than its ability to resolve, and be an on-going part of, problem-free value theory. To defend this position, I will demonstrate the resources available for making the counterintuitive consequences seem less counterintuitive. This

will involve two steps. First, I will introduce a dispositional account of grounds for uncertainty, and second, I will endorse a pluralistic notion of normative communities.

When I introduced the notion of cause of uncertainty in Chapter 3, I meant it literally. An agent with cause for uncertainty has literally been caused to be uncertain. But how does such actual uncertainty relate to grounds for uncertainty? I define grounds for uncertainty as follows: If, holding everything else in your present condition fixed, your attending to (experiencing/noticing/discovering) information i will cause you uncertainty, then i is grounds for uncertainty, whether you attend to i or not. For example, if you were to discover that all of the paint in your house has trace amounts of lead, this would cause you uncertainty about the safety of your house. Thus, the existence of the lead is grounds for you to be uncertain about the safety of your house, even though you are not in fact uncertain about the safety of your house because you do not know about the lead. The existence of unnoticed grounds for uncertainty is exactly why the habit of checking for uncertainty, discussed in §3.3.4, is useful.

With this dispositional understanding of grounds for uncertainty, we can accommodate a great deal of the circumstances we are likely to judge cases of mistaken certainty and uncertainty. But notice, these cases cause no problem for the evidential account of normative authority, because it is built into the theory that the authority of a given value adjustment resides not only in its ability to resolve problems currently, but in its ability to be part of an on-going problem-free value theory. Thus, the vindication of a given value adjustment is relative not just to its capacity to resolve actual uncertainty, but also to its capacity to resolve, or at least be part of a solution to, potential uncertainty.

And, given the plethora of unnoticed, but available, information, there is a great deal of potential uncertainty lurking around every corner.

However, the dispositional account of grounds for uncertainty will not explain away all cases that we might want to call cases of mistaken certainty and uncertainty. There are two ways in which an agent might fail to appreciate grounds for uncertainty. First, she might simply overlook, i.e., fail to attend to, available information that would cause her uncertainty if she attended to it, in which case uncertainty and deliberation will not be triggered. Second, even when attending to grounds for uncertainty, an individual might be caused no uncertainty. Call the first type of failure to appreciate grounds for uncertainty a case of overlooking and the second type of failure a case of unresponsiveness.[100]

My approach to the second type of cases is to deny that they are in fact cases of mistaken certainty, and instead appeal to a response-relative definition of epistemic and normative communities. To take a clear, though extreme example, the psychopath is not a member of our normative community precisely because the perception of another's suffering causes him no pause.[101] Insofar as the normative uncertainties he faces are dramatically different than the ones we face, we should not expect the value judgments he makes and tests in experience to be binding on us. Of course, there is still a question

[100] Nomy Arpaly recognizes a similar distinction, which she calls the distinction between "deficiency of perception" and "deficiency of motivation." 2003. (83) She writes, "Many people have wondered if their spouses don't see the dust on the floor or don't care that it is there, and the diagnosis is hard because there is often a combination: a person who does not care about cleanliness tends to be less trained at spotting dust, and a person who is not good at spotting dust may find it hard to be motivated to learn to see it and add a new worry to his life." (83)

[101] Of course, there is the chance that psychopaths simply do not perceive the blatant behavioral signs of suffering as signs of suffering. The problem then would be that because there is no way for them to attend to the suffering of others, they have no access to the fact that another being is suffering, in which case we cannot judge whether their failure to respond is due to overlooking or unresponsiveness.

of how we should, practically speaking, deal with a psychopath, just as there is a question of how we should deal with a lion, a question that is not answered by characterizing his lack of responsiveness to evidence as differences rather than mistakes.

One key objection to this strategy might be the claim that it narrows one's normative community much too far, or, worse, that it is "shockingly cultlike."[102] The worry here is that we will take the lesson of the psychopath too far. Does this theory sanction us to simply excise from our normative community any individual who has an opinion different from our own?

There are two positions from which this objection may be raised, and I will take them in turn. The first is the position that, whatever normative theory we settle upon, we must take it as a constraint upon that theory that all adult humans are equal and reciprocal members of a shared normative community. I see no reason beyond sheer dogmatism to insist that we treat this result as a constraint upon our normative theory. It is a live question, I believe, whether and in virtue of what we humans are members of the same normative community.

However, there is a second, non-dogmatic position from which to resist the division of the normative community along lines of shared responsiveness. The success of pluralistic communities like ours have demonstrated that in spite of vast differences in responsiveness, differences that sustain, presumably, our society's widely varying metaphysical and ethical beliefs, humans are able to forge workable political and social communities, to articulate somewhat shared conceptions of the good, and to agree upon principles of justice. This suggests that excising someone entirely from the normative

[102] I thank Elizabeth Anderson for making this objection clear.

community because of a difference in responsiveness to a particular set of evidence is a dramatic overreaction.

With this I agree. But nothing I have suggested so far endorses this picture of dramatic excision. To make this clear, allow me to highlight two points. First, because differences in opinion can result from either overlooking information or unresponsiveness to information, differences in opinion alone should never be treated as definitive evidence of the latter. In fact, because of the high cost of subdividing the normative community, the presumption that differences in theory and opinion are predicated on differences in attending, as opposed to differences in responsiveness, is always warranted. One should make sure that disagreeing parties are attending to all the same information (under the same description) before one writes a disagreement off as irreconcilable.

Second, there is no reason to assume that normative communities are mutually exclusive, all-or-nothing affairs. The conditions that make us the type of beings that have problems do not guarantee that we share any specific problems. And yet, as a matter of fact, we do share a great deal of our problems. This is because we come from similar biological stock and we are tethered together in our physical, social, linguistic, and cultural environments. These facts make it likely that we will have similar epistemic, evaluative, motivational, and phenomenological reactions to the world we encounter. These dispositions, which are the fodder for the problems we face, the problems we must figure out how to overcome via deliberation, define the boundaries of our normative communities.

In spite of these similarities, there are also variations in our responses to the world. Consider, for example, that alcoholics are well served by banding together to

devise strategies for overcoming their problematic addictions. The norms that serve that purpose—such as absolute abstention from alcohol—are not binding on non-alcoholics. So too, the normative strategies which are the fruitful products of problem-solving efforts of any particular sub-group of problem-havers are not binding on those who do not have those problems.

And so, I propose, the degree to which a given individual values exercise, family, money, art, sex, prestige, power, adventure, and intellectual engagement rightly depends on what problems would be caused by giving up such values. It is, I believe, unreasonable to expect that this measure will be the same for all humans in all situations. Thus, each person belongs to a variety of overlapping normative sub-communities, and each of these sub-communities will be bound by norms specific to the problems their members face. But sub-communities are not cults. An honest attempt to discover which sub-communities you belong to, and to understand the values you develop as relative to those communities, is neither cult-like nor shocking. It is, rather, another way of describing the familiar human quest for belonging.

Further, if there is one normative community to which all humans and only humans belong, it will be in virtue of the fact that we all share, or are disposed to share, the same substantive problem or problems, not in virtue of the fact that we are all problem-havers, nor, for that matter, in virtue of the fact that we are all reflective beings.

In any case, whether or not there is a single normative community to which we all belong, there is certainly a vast amount of commonality among the problems we are disposed to face. This contingent fact about our shared predicament, in conjunction with our capacity to understand and communicate about the commonality among our

problems, creates the possibility of a shared project of value assessment, as well as the possibility of collaboration and cooperation in revising and innovating new value structures. We can learn from others what matters and why. We can share solutions with others, and they can share their solutions with us.

The evidential theory gives us a framework for understanding how and why it is possible to challenge the legitimacy of a given value set, just as the evidential theory of science gives us a framework for understanding how and why it is possible to challenge the legitimacy of certain theoretical beliefs. Namely, attention to new information can cause uncertainty about our values, and this uncertainty will make it impossible for us to continue without revising those values.

But this framework does not itself deliver any verdicts as to which particular values have authority. For this, we must turn to life itself, or, within philosophy, to applied ethics. When it comes to our current plans, there is no shortage of unnoticed grounds for uncertainty. Applied ethicists are like experimental scientists. They advance the field of ethics by identifying and bringing to our attention the ways in which our current moral theories are insufficient, and they help formulate and articulate new principles and values to overcome these insufficiencies.

4.2.2 A Note on normative discourse

During open-ended reasoning, diagnostic reason often leads us to discover that some, or many, of our activated goals are supported not just by unreflective desires or impulses, but by a history of problem-solving episodes. In other words, many of our goals have been explicitly adopted in order to solve previous prospective problems. Or, they have thrived in prospection because, though accidentally adopted, they nevertheless

harmonize our motivations effectively. In such cases, deactivating these goals will not simply involve ignoring the demands of a persistent desire, but rather, will involve opening or re-opening the prospective problems they hold at bay.

For example, accepting the ideal of democracy creates many problems for our society. Holding elections and providing education to a population is expensive. It is also harder to implement needed changes when everyone has a say in whether and how those changes are implemented. These problems could be avoided if we simply abandoned the ideal of democracy. The reason that we continue to champion democracy in the face of these challenges is not just that we want to have a democracy, nor that we find democracy inherently satisfying (though we might). Rather, democracy is itself a solution to numerous other problems. Hence, giving up on democracy would require not just frustrating a desire, but rather, reopening additional problems, many of which are worse than, say, figuring out how to fund an election.

Values like this, values that play problem-solving and problem-preventing roles, provide important structural support to our prospective economies. When someone challenges the authority of these norms, we have something to say in their defense. This defense will involve many other values and beliefs, and an appeal to experience and predicted consequences. All of these are the proper subjects of normative discourse. Are these values as supportive as we take them to be? Are there other values that would provide better support? Is the support they currently provide predicated on any mistaken beliefs about the world? And so forth.

We can interpret moral discourse on topics like same-sex marriage, abortion, animal rights, euthanasia, prison reform, affirmative action, medical ethics, justice in

distribution, religious freedom, etc. as engaged in exactly this kind of questioning. Part of what we might discover in the context of such discourse is that not everyone involved in the discourse is disposed to exactly the same types of problems as we are. But, there is more similarity between us than difference, and so moral discourse can yield fruitful understandings of shared problems, and help us craft more robust and enduring solutions to those problems.

Chapter Five

Deliberation and Responsibility

In this chapter, I will argue that the prospectivist account of motivated behavior, in combination with a diagnostic conception of practical reason, provides an attractive answer to the question of whether, and in what regard, human choice and action is free. To avoid confusion, I want to be clear from the outset that I am not seeking to defend the existence of human free will as it has been traditionally defined, i.e. as categorical freedom of choice. An agent possesses categorical freedom of choice only if, for any decision d that an agent has made, she could have chosen differently under exactly the same circumstances. However, I hope to show that prospectivism's conception of intelligent deliberation answers to most, if not all, of the same data and desires to which free will supposedly answers. If I succeed, I will leave the reader with the following question: Given that prospectivism's understanding of intelligent deliberation is both compatible with our general scientific understanding of the world[103] and capable of satisfying the same data and desires that categorical free will supposedly satisfies, why worry about whether we have the latter?

I proceed in the following manner: First, I investigate the reasons people are typically compelled to believe in categorical free will. Second, I explain prospectivism's conception of intelligent deliberation. Third, I show that prospectivism's notion of

[103] This includes determinism for mid-sized objects, but allows for indeterminacies at the quantum level.

intelligent deliberation is compatible with determinism[104] by testing the theory against a common incompatibilist attack. In the process, I demonstrate why prospectivism's brand of soft determinism is more promising than other versions on the table. Finally, I explore the ways in which prospectivism's conception of intelligent deliberation answers to the phenomenology and desires that typically lead people to believe in categorical free will.

5.1 Reasons for Believing in Free Will

There are three primary reasons that people are tempted to believe in free will. The first is a set of phenomenological, or introspective, data that requires explanation. The second is a strong desire, shared by many, for a particular type of agential control. And the third is the fear that, in surrendering the notion of free will, we would thereby be forced to surrender our practices of holding people morally responsible. I will take these in turn.

Consider the way it feels to be in the midst of a difficult choice. Imagine, for instance, choosing between sending your child to a private school—where she will get a great education, but which also costs more than you can comfortably afford and involves supporting an elitism you are uncomfortable with—and sending her to a public school, where she will have far fewer educational resources, but which will cost very little and involve adhering to your fundamental belief that education should not be a matter of privilege. Some would suggest that the outcome of your deliberation is a matter of the blind forces of physics, but of course, it does not feel that way to you. You are aware of

[104] For the purposes of this chapter, I will use the term determinism to denote the position that human minds are subject to causal laws just like everything else in the universe. This use allows for the possibility that there may be some probabilistic or indeterministic physical laws at the quantum level.

all your relevant desires—your desire for your child's happy and healthy future, your desire to be fiscally responsible, your desire not to be an elitist, etc.—and yet your full awareness of these desires, and even your initial attempts to weigh their respective values, is not enough to determine what you should or will do. If it were, proceeding would be simple. Instead, proceeding involves making an extraordinarily difficult choice. It seems that the world, your body in particular, is waiting for instructions and that nothing but you, in particular your mind, is going to provide those instructions (especially not the laws of physics and some contingent facts about your environment).

To be clear, there are four relevant phenomenological pieces of data here. 1) It seems to you that you, in particular your choice of a course of action, will be the proximate cause of your actions. 2) It seems to you that unless you actively engage in a certain type of mental activity, no choice will be made at all, or in other words, that your deliberation is inherently active rather than passive. 3) It seems to you that when you do eventually make a choice, you will be responsible for that choice, because right now, your decision is the only thing in the world that is going to determine the matter one way or the other. 4) It seems to you that you are in a position to freely decide among a variety of genuinely available options. This includes the appearance that nothing—internal or external—is forcing you to choose one option over another.

According to the traditional free will dialectic, the determinist[105] will maintain that such phenomenological data is not veridical. In fact, determinists are often eager to

[105] Again, for the purposes of this chapter, I will use the term determinism to denote the position that human minds are not categorically free and that they are subject to causal laws just like everything else in the universe. Along the same lines, I use the term determinist to denote a generic opponent of libertarianism, someone who believes that the activity of the human mind is governed by the same laws that govern physical events. As such, a determinist need not believe

point to numerous instances in which phenomenological data has led us astray (e.g. we thought the earth was flat because it seemed that way to us, but we were wrong) and simply insist that free will is just one more phenomenological illusion. However, I am sympathetic with those who feel this dismissal is often too careless. Yes, the phenomenological data sometimes leads us astray; but also, sometimes it does not. And though the evidence for determinism is generally very strong, it is at its weakest when it comes to understanding the causal relationship between mental events and physical events. In fact, no one has ever witnessed the generation of a choice, except from the first-person perspective. And from that perspective the generation of the choice seems to involve a certain amount of freedom (at least in many instances). Therefore, this data deserves closer attention. What is the content of our phenomenological experience when we are deliberating and is it indeed incompatible with determinism?

The second reason we may be tempted to believe in free will is a simple one: we want to be categorically free. There is a large emotional payoff in adopting the belief that, as agents, we are free from the demands of the causal nexus, that we stand somehow apart from the universe's molecular tide, and that we engage with the physical world only in accordance with our self-contained will. In short, we like to believe we are self-determined and in control of our actions.

Of course, wanting a certain theory to be true does not constitute evidence of its truth. However, our beliefs about our freedom do more than make us feel good, they also have tremendous pragmatic value.[106] Thus, when presented with determinism, a theory

---

in strict determinism, but rather can allow for probabilistic or indeterministic laws at the quantum level.

[106] Something has pragmatic value to the degree that it helps us navigate our world and make sense of it and our reactions to it. The more pragmatic value a belief has to us, the less we would

that has its own merits and epistemic support, the proper attitude is not the hasty abandonment of our current beliefs, but a cautious attempt to reconcile them with determinism. The appropriate question in this case is: Even if determinism is true, am I still warranted in retaining a belief that has enormous pragmatic value—am I warranted in regarding myself as self-determined and in control? Perhaps it will turn out that the pragmatic value of belief in freedom can be realized by belief in some type of control short of categorical free will.

This leads to the final reason for believing in categorical freedom of choice: the fear that in surrendering free will, we will have to give up the possibility of moral responsibility. Some have suggested that if in fact human choice is not categorically free, then it is completely nonsensical, and perhaps cruel, to blame an agent for the unsavory things she has done. Consider, if everything I am doing right now has been determined, necessitated, by forces beyond my control, it seems intuitively unjust to hold me responsible for these actions, or at least as unjust as it would be to hold someone responsible for falling down because they have been physically thrown to the floor by an earthquake.

Again, our commitment to the existence and importance of moral responsibility provides a pragmatic reason for belief in free will, and the determinist's response is obvious: Even if we currently hold people morally responsible for their actions, and even if we cannot imagine what it would be like to continue without these practices, these facts alone are not proof that blaming and praising people for their actions is not cruel or absurd. In fact, it may be that asking who is morally responsible for the destruction of

---

know how to carry on without it. A great deal of Dewey's work in epistemology has centered around articulating this concept.

the World Trade Center in New York makes about as much sense as asking who is morally responsible for the tsunami that killed hundreds of thousands of innocent civilians in the Indian Ocean two years later. If both events followed inevitably from the blind forces of physics, the correct answer in both cases is no one.

And yet even the determinist must admit that there is an intuitive difference between the two cases. The latter invokes only a feeling of sorrow and the hope that we may be able to prevent, through technological and educational advances, such tragedy in the future. The former evokes, in addition to these feelings, a sense of moral outrage. It may be the case that the differences in our emotional and rhetorical responses to the terrorist attacks of September 11, 2001 and the tsunami of December 26, 2004 do not track any actual metaphysical difference between the two events. But surely we should not jump to this conclusion just because we think determinism is true.

Sequences of events initiated by human minds are, and have always been, especially interesting to us, probably for a good evolutionary reason. And even if it turns out that human deliberation is not categorically free, and hence does not initiate chains of events in the strictest sense of 'initiate', that hardly seems reason enough to stop taking a particular interest in the choices we humans make, the reasons we make them, and the events that follow from these choices. Before we discard the possibility that humans may be morally responsible for their actions, we may want to reexamine exactly why it is important to us, and whether some other conception of human freedom might support our moral rhetoric and practices, or perhaps recommend even better practices.

Before I go any further, I want to briefly note a few reasons for believing in categorical free will that I will not attempt to accommodate. The first is the dominant

contemporary folk psychology of our culture. It is commonplace for people to claim that humans have free will. And this fact in itself provides a certain amount of support for believing in free will. Secondly, there are theological reasons for believing in free will. For instance, free will may be the best defense against the problem of evil; free will may be a prerequisite for admittance into heaven; or free will may be an intrinsic feature of souls. However, these are not my starting points, and I see no reason to make them so. Thus, while I hope to show that prospectivism's notion of intelligent deliberation is compatible with our phenomenological data and many of our desires and practices, I cannot promise that his theory is compatible with all our deeply held beliefs.

To summarize, I have offered three reasons people are typically compelled to believe in categorical freedom of choice: phenomenological data, the desire to be self-determined, and the need to continue to regard ourselves and others as moral agents—as worthy of moral praise and deserving of moral blame. All three need to be taken seriously. What I hope to show soon is that prospectivism's non-categorical notion of rational control is compatible with both determinism and our phenomenological data; that it satisfies our desires for self-determination; and that it is capable of supporting and making sense of our practices of deliberation, of attributing responsibility, of holding people accountable for their actions, and of morally praising and blaming others and ourselves.

5.2 Intelligent Deliberation

I would like to propose that an agent is free in a sense that answers to the above considerations if and when she employs her capacity for intelligent deliberation. Intelligent deliberation is deliberation that allows an agent's understanding of the

problem that has necessitated her choice—an understanding that can be more or less accurate and more or less useful—to help her achieve that choice.

On this picture freedom does not involve the ability to initiate chains of events, but rather to reorganize and redirect chains of events in a manner that is meaningful and promising to the agent. Freedom involves being so designed that thoughts about the future, in particular thoughts about the predicted consequences and expressive meanings of imagined actions as they relate to a given deliberative impasse, can serve as both stimuli and sources of significance. Shortly, I will discuss whether intelligent deliberation is an adequate definition of freedom. First, however, I want to establish that freedom defined in terms of intelligent deliberation is in fact compatible with determinism.

5.3 Intelligent Deliberation and Determinism

Propsectivism's account of freedom as intelligent deliberation faces potential challenges from two directions. First, there are philosophers like A.J. Ayer and Walter Stace who agree that humans are free and that this freedom is compatible with determinism (a position known as soft determinism), but who offer rival accounts of what human freedom consists in. Second, there are philosophers, such as Richard Taylor,[107] who argue that no type of freedom is compatible with determinism. We can deal with both of these challenges at once by demonstrating that prospectivism can survive Taylor's incompatibilist attack while Ayer and Stace cannot.

[107] I have chosen to use Taylor as my particular incompatibilist foe, but incompatibilism has a wide canopy, including libertarians, who believe we have free will and determinism is false (for example C.A. Campbell, Peter van Inwagen, and Robert Kane), and hard determinists who believe that determinism is true and there is no meaningful sense of freedom that is compatible with determinism (such as B.F. Skinner).

In his article "Freedom and Determinism" Taylor offers the following thought experiment:

> …suppose that an ingenious physiologist can induce in me any volition he pleases, simply by pushing various buttons on an instrument…All the volitions that I have in that situation are, accordingly, precisely the ones he gives me. By pushing one button, he evokes in me the volition to raise my hand; and my hand being unimpeded, rises in response to that volition…We can even suppose that the physiologist puts a rifle in my hands, aims it at some passer-by, and then, by pushing the proper button, evokes in me the volition to squeeze my finger against the trigger, where upon the passer-by falls dead…This is the description of a man who is acting in accordance with his inner volitions, a man whose body is unimpeded and unconstrained in its motions, these motions being the effects of those inner states. It is hardly the description of a free and responsible agent. It is the perfect description of a puppet.[108]

The argument in this passage can be made explicit as follows:

P1) An agent whose mental states are being manipulated by a physiologist in the manner described above is not free. (True by strong intuition.)

P2) The reason she is not free is because the mental states that compose her will are being determined entirely by forces beyond her control. (True by inference to the best explanation.)[109]

P3) If determinism is true, then in all instances the mental states that compose an agent's will are determined entirely by forces beyond her control, namely, the laws of physics and the initial conditions of the universe. (True by definition of determinism.)

P4) Determinism is true. (True by supposition.)

---

[108] Taylor, Richard. 1963/1999. (445)

[109] Taylor attempts to block an anticipated rival explanation of the agent's lack of freedom with the following passage: "Whether a desire which causes my body to behave in a certain way is inflicted upon me by another person…or derived from hereditary factors, or indeed from anything at all, matters not the least" (445). His point here is that it is not the fact that another person is controlling the agent that deprives the agent of freedom. I tend to agree with this claim—it is hard to draw a principled distinction between being forced to do something by a person and being forced to do it by anything else. Of course, there may be something extra-bad about being forced to do something by another person, just like it might feel worse to have your car stolen than to have it washed away in a flood. Either way you lack the car because something removed it from your possession, but the fact that this something was another person might add insult to the injury.

C1) In all instances, the mental states that compose an agent's will are determined by forces beyond her control.

C2) In all instances an agent is not free.

To defeat this argument, a soft determinist must disprove one of the first three premises. The most promising way to do this—or at least, the one that is most often tried—is to offer an account of freedom that makes the second premise false. What the soft determinist is looking for is an alternative explanation for why the agent who is being controlled by the physiologist is not free: in particular, an explanation that does not hold true of persons in general.

In his article "The Problem of Free Will" Stace defines human freedom as follows:

> …it is obvious that all those actions of men which we should ordinarily attribute to the exercise of their free will, or of which we should say that they freely chose to do them, are in fact actions which have been caused by their own desires, wishes, thoughts, emotions, impulses, or other psychological states.[110]

Stace is correct that if human freedom is defined as such, it is compatible with determinism. His alternative to Taylor's second premise would be this:

$P2_S$) The reason the agent is not free is because the mental states that compose her will are not her own.

This response, however, begs the question: Under which circumstances are an agent's mental states legitimately her own? In the face of Taylor's challenge, Stace must justify the assertion that mental states entirely determined by ingenious physiologists are not an agent's own mental states, while still maintaining that mental states entirely determined by the laws of physics and the initial conditions of the universe are the agent's own. This

[110] Stace, Walter. 1952/1999. (417)

may not be an impossible task for Stace to execute, but as the dialectic stands, the ball is in his court.

A second attempt at a formulation of soft determinism can be found in the article, "Freedom and Necessity" by A.J. Ayer. Ayer articulates his understanding of freedom in the following passage:

> [To say I am free] is to say, first, that I should have acted otherwise if I had so chosen; secondly, that my action was voluntary in the sense in which the actions, say, of the kleptomaniac are not; and thirdly, that nobody compelled me to choose as I did.[111]

The third criterion of Ayer's definition of freedom (that nobody, that is, no other person, is determining an agent's choices) seems, at least at first glance, to give him an easy refutation of second premise of Taylor's attack. Ayer's alternative to Taylor's second premise would look like this:

> $P2_A$) The reason the agent is not free is because the mental states that compose her will, and hence her choices, are being determined by somebody other than herself.

And yet, this again begs the question against Taylor. To see how, we need to step back and look more closely at the method Ayer uses to arrive at his conclusion.

Ayer begins from two observations: that freedom is usually defined in contrast to constraint, and that causal determination is not necessarily equivalent to constraint. Therefore, it follows that freedom does not necessarily require lack of causal determination. Another way of putting his point is this: compatibilism is possible if freedom is defined, as it is and should be, in terms of lack of constraint. Next Ayer poses the apt question: "In what circumstances can I legitimately be said to be constrained?"[112] The supposition is that if the answer to this question involves a more limited set of

[111] Ayer, A.J. 1969/1999. (412)
[112] Ibid. (411)

circumstances than the answer to the question "In what circumstances can I legitimately be said to be causally determined?" then Ayer has proven his case.

The question-begging aspect of Ayer's approach is how he proceeds to define constraint, namely, by establishing a list of circumstances that intuitively involve constraint. For example, he writes, "an obvious instance is the case in which I am compelled by another person to do what he wants."[113] In a similar manner, he cites hypnosis, habits of obedience, physical constraint, psychological duress, and neurosis as obvious cases of constraint. There are two points to be made against Ayer here. First, "legitimate" and "obvious" are not synonyms. So to answer the question of which cases are legitimately described as involving constraint by investigating which cases obviously involve constraint seems misguided. Second, while Ayer creates an intuitive list of circumstances that involve constraint, he does nothing to justify his omissions from this list, most importantly, his omission of causal determination as a form of constraint.

Remember, it is exactly the fear that if determinism is true, all our actions are in some legitimate sense constrained (by physical law and prior conditions) that began the free will debate in the first place. Thus, if Ayer hopes to defuse this fear, he has to do more than claim that the laws of physics are not what we normally think of when we think of things that constrain. He has to give us some positive reason not to think of them as constraints.

To return to our starting point, Taylor can block Ayer's response by demanding an answer to the question: Why does being determined to act in a certain way by another person count as being constrained, while being determined to act in a certain way by laws

[113] Ibid. (411)

of physics does not count as being constrained? If Ayer cannot answer this question, he has not made any headway against Taylor.

Intelligent deliberation fares better against Taylor than Ayer and Stace do. Prospectivism begins from the Deweyan starting point that an agent's habits, desires, ambitions, aversions, and goals are substantially influenced by her social environment. Everyday, the behavior of the people in our lives, from marketers to our closest friends, affect what we believe and what we desire. Taylor's ingenious physiologist represents just an extreme, and particularly invasive, case of something that happens all the time. Because the notion of intelligent deliberation is built against a backdrop of such external forces of influence, we may expect it to be better equipped to accommodate Taylor-style objections than rivaling soft determinist theories.

In the face of Taylor's challenge, prospectivism's conception of freedom needs a two-pronged approach. For, depending on the nuances of how the physiologist in Taylor's thought experiment manipulates the agent's mind, prospectivism will charge either that Taylor's second premise is false, i.e. that the agent is unfree not because her actions are being determined by forces outside her control but for some other reason, or that Taylor's first premise is false, i.e., that the agent is in fact free, despite our initial intuitions on the matter. I'll start with the rejection of the second premise.

The agent in Taylor's thought experiment appears to be incapable of intelligent deliberation. In fact, the way the agent's behavior is described, it sounds as if the physiologist has managed to turn her into a child-like agent, someone who unreflectively acts on impulse. Prospectivism accepts that children possess a great deal less freedom than "normal" adults precisely because they cannot deliberate intelligently. In order to

deliberate intelligently, an agent must have both the capacity to be made uncertain, and the capacity to resolve this uncertainty by way of her own reflective diagnostic reason. Those who act entirely on impulse are such that, because they never have the opportunity to employ diagnostic reason in the face of uncertainty, never intelligently deliberate. Thus, prospectivism can replace Taylor's second premise

> P2) The reason an agent is not free is because the mental states that compose her will are being determined entirely by forces beyond her control.

with something like this:

> $P2_D$) The reason the agent is not free is because the physiologist has deprived her of the capacity to deliberate intelligently.

Because this condition—the lack of the capacity to deliberate intelligently—is not universal to all agents even if determinism is true, such a move blocks the deduction to the conclusion that all agents are not free.

In response to these moves, Taylor might claim that I have missed his point. If a physiologist actually had complete control over an agent's mental states, he would not need to turn the agent into an impulsive agent in order evoke in her the desired intentions. Rather, he could simply set up an agent's mentalscape so that it simulated, perfectly, the mentalscape of someone who would in fact *intelligently choose* to do the physiologist's bidding. Suppose, for example, the physiologist wanted the agent to surrender her diamond ring. Could he not, by manipulating the agent's mind, make her intelligently choose to do so? The answer to this question is less than straightforward.

First, assuming prospectivism entails that delusion is a restriction on freedom[114], the physiologist cannot make the agent <u>intelligently</u> choose to give up her ring simply by

[114] An agent who is drastically misinformed about the world, and specifically about the consequences of his actions, cannot deliberate intelligently. Coordinating our habits with our

manipulating her beliefs, for example, by making her believe her ring actually belongs to the physiologist. So again, if this is how the physiologist is manipulating the agent's mentalscape, then the second premise of Taylor's argument is false. But assume instead that the physiologist overhauls the agent's entire character, including all of her background proclivities, impulses, and desires. Suppose the physiologist designs the agent's mind so that whenever she is confronted with the opportunity to aid scientists, she derives meaning and satisfaction from deciding to do so. Under these circumstances, when the agent correctly recognizes the physiologist is a scientist, and correctly recognizes that giving the physiologist her ring will aid him, and correctly recognizes that if she gives the physiologist the ring, she will still have plenty of money to meet her needs and desires, her decision to hand over her ring seems to fit the description of intelligent choice. (It seems analogous at least to what prospectivism would consider a free choice to donate money to a charity, or to give money to your children, etc.).

This brings us to the second prong of prospectivism's defense. Prospectivism's understanding of freedom as the exercise of intelligent deliberation commits it to the conclusion that, in this final permutation of the thought experiment, the agent is free. This clashes with our initial intuition on the matter, the point at which we nodded our heads in agreement with Taylor's first premise: an agent whose mental states are being manipulated by a physiologist in the manner described above is clearly not free. Prospectivism's best move in this position is to claim that our intuitions are wrong, and to attempt to explain them away. We can do so as follows.

---

environments requires that our habits respond to our *actual* environments. Any habits that clash with the actual environment will, if our minds are properly functioning, be quickly revised. Thus, a requirement for veridicality of belief is built into prospectivism.

First, it is important to point out that the thought experiment, as it was originally conveyed, gave the impression that the agent had been almost literally turned into a puppet. The physiologist seemed to be giving direct orders to the agent's motor-cortex (in the form of "volitions"), thus cutting the agent's deliberative apparatus out of the loop. When we whole-heartedly agreed that such an agent lacks freedom, it is not clear that our intuitions stemmed from our belief that the agent's volitions were determined by forces beyond her control. It is equally plausible that our intuitions stemmed instead from our belief that her volitions were not formed via the proper deliberative process, or in other words, from our sense that the agent was not given the opportunity to reflect intelligently on the impulses and habits that were moving her to act. Once the thought experiment is redescribed to include the fact that the volition motivating the agent to act was the result of a conscious, desire-harmonizing session of deliberation carried out in the mind of the agent, it seems that our intuitions that she lacks freedom become a lot less overwhelming.

Furthermore, to the extent that these intuitions still manifest themselves, we can explain them away in terms of a confusion between violation of and restriction upon freedom. In other words, our misleading intuitions stem from our acknowledgement that the agent has been violated, that she has been forcibly changed in a way that, had she been asked, she would not have accepted. We are certain the agent is a victim; we are repulsed that the physiologist is so blatantly using her as a mere means to his ends. We believe strongly that the agent as she was prior to manipulation would not now approve of her decision to hand her ring over to the physiologist. This, in turn, leads to the sense

that she is choosing to do so against her will, and this is the point at which our intuitions about her lack of freedom kick in.

But we are wrong that she is choosing against her will. Tragic and unfair as it might be, the agent has a new will now. And if we can keep this fact in mind, our intuitions that her choice is not free can be defused. Thus, prospectivism's final response to Taylor can be summarized as follows: By manipulating an agent's mental states, either the physiologist deprives an agent of freedom by limiting her capacity for intelligent deliberation, or the physiologist does not restrict the agent's freedom at all, but rather creates a new problem for the agent, which the agent then intelligently resolves.[115] Either way, the argument for incompatibilism fails.

Before I move on, I want to consider one more way that the physiologist could possibly "force" an agent to "freely decide" to hand over her ring: He could ditch his fancy technology, leaving the agent's current mentalscape untouched, and instead grab a gun and threaten to kill her unless she surrenders the ring. If the agent decides to hand over her ring to avoid death, has she acted freely? The standard intuition in reaction to this question is "no." And certainly for legal purposes we want to say that the robber forced her to give him her ring.

Is it a problem, then, that prospectivism commits us to the position that in such circumstances, as long as the agent has openly considered all her available options under their various descriptions and in light of their various consequences, she has chosen freely? This question points to an underdeveloped portion the of prospectivist picture: In some cases an agent's options can, objectively speaking, be so limited that any choice she

[115] Of course, an accurate understanding of her problem would include the fact that it was caused by the manipulation of her mental states by the physiologist. This might be alone sufficient to make an exception in this case to her desire to help scientists.

makes is going to involve ignoring or suppressing a value that she takes to be authoratative. (For a poignant example, think of Sophie's choice.)

Stace is happy to relegate these cases to a gray zone of 'borderline cases', to claim that in some instances it is simply not clear whether we are free.[116] But prospectivism can offer a more sophisticated response. In particular, the prospectivist can distinguish between two distinct types of freedom. The first, deliberative freedom, is measured in accordance with the degree to which an agent exercises an intelligent response to the objective contours of a deliberative impasse, whatever those contours happen to be. The second, freedom of opportunity, can be measured in accordance with the resources available to the agent during her impasse. [117]

To make the difference between these two types of freedom clear, consider the following case. Suppose there are two agents, one in a room with three doors and one in a room with two doors. Further, assume the both have to pick a door to pass through, and both agents can pass through only one door at a time. (Imagine that the agents know something about where the doors lead, etc.) Given their respective deliberative impasses, the agents could be equally open and creative when rehearsing in imagination their respective options, and equally thorough when weighing these options. If they are so equal, then the agents share the same degree of deliberative freedom. But there is also a sense in which the agent in the room with three doors has more to work with than the

[116] In discussing an armed robbery case, Stace writes: "Aristotle…admitted that there are what he called 'mixed' or borderline cases in which it is difficult to know whether we ought to call the acts free or compelled. In the case under discussion, though no actual force was used, the gun at your forehead so nearly approximated to actual force that we tend to say the case was one of compulsion. It is a borderline case." (Stace, Walter. 1952/1999. (417))

[117] I have intentionally avoided relying on Locke's distinction between free will and freedom of action here to avoid some of the problems that come with it. However, I believe his distinction and the one I am introducing here are motivated by the same intuitions and theoretical goals.

agent in the room with two doors. When the agent in the three-door room imagines various courses of action, there are three different general descriptions these actions fall under, corresponding to three general sets of consequences she has to choose among; the other agent has only two. Hence, the three-door agent has a greater freedom of opportunity than the agent in the room with two doors has.

The important point is that there are two ways to restrict an agent's freedom. One is to impair her capacity for effective intelligent deliberation, thus restricting her deliberative freedom. The second is to manipulate the objective circumstances of her deliberative impasse so that she has very few resources to work with, thus restricting her freedom of opportunity. In the case of the armed robbery, our intuition that the agent's choice to surrender her ring is not free can thus be explained in terms of her lack of freedom of opportunity. Under such circumstances, all the options available to her entail only two possible salient consequences: the loss of her ring or the loss of her life. Her failure to generate in imagination any possible courses of action that might lead to different, more appealing, circumstances is not due to a failure of imagination or intelligence, but rather, to facts about the environment external to her deliberative processes, facts which are the subject matter of her diagnostic deliberative processes.

Both deliberative freedom and freedom of opportunity are important to prospectivism. To increase an agent's freedom of opportunity is to increase the potential range of her deliberative freedom (in the same sense that giving a talented painter additional colors to work with can increase the range of what she can produce, without directly affecting her abilities as a painter). Second, and perhaps most importantly, both types of freedom are compatible with determinism. Freedom of opportunity is measured

in terms of the available outcomes[118] to an agent during a deliberative impasse, and deliberative freedom is measured in terms of how intelligently an agent can conceive of the contours of her impasse in order to overcome her impasse, which depends on the talents of her diagnostic reason.[119] The truth of determinism diminishes neither of these freedoms.

At this juncture, it seems that prospectivism can undo Taylor's incompatibilist argument, and in the process prove that prospectivism's version of freedom as the exercise of intelligent deliberation is superior to the soft determinism offered by Ayer and Stace. However, there is one last charge that incompatibilists (and perhaps compatibilists as well) can make against prospectivism, a claim that might be phrased as such: "Prospectivism, in attempting to maintain that freedom is compatible with determinism, has twisted the meaning of the word 'freedom' to such a degree that it no longer captures what we care about, say, when we desire for ourselves to be free, and that it can no longer do the work we believe freedom can and should do to justify or moral practices and explain our phenomenological experience."

Thus, I now turn back to the question I began the chapter with—why is freedom important to us? In the following section I hope to establish that deliberative freedom, in

---

[118] An avid incompatibilist might try to challenge the notion that any outcomes are actually 'available' to an agent except for those that she is determined to pursue by the laws of physics. But if this is the case, an agent in the room with two doors has precisely as many options available for exiting the room as an agent in a room with one door—namely, one. Settling upon a definition of availability that made this claim true would involve drastically overhauling, abandoning even, all of our modal concepts.

[119] Again, an avid incompatibilist might object: "But if that's your definition of freedom then possessing freedom is a matter of mere chance." The response to this objection is yes and no. Yes, like possessing beauty, or being born into a large fortune, or being born human, possessing the capacity for intelligent deliberation is to some extent a matter of luck. However, if an agent happens by chance to come to care about possessing the capacity for intelligent deliberation (i.e. if she happens to come to care about being free), then she can do something about it, at least to the extent allowed by her physiology, because her desire for freedom can then figure into her choices.

combination with adequate freedom of opportunity, is exactly the type of freedom that we do and should care about.

5.4 Is Intelligent Deliberation Enough for Us?

According to the picture I have been defending here, an agent is free if 1) she has the capacity to deliberate intelligently, 2) she exercises that capacity in such a way as to effectively direct action, and 3) she is in a situation that affords a certain minimum amount of freedom of opportunity. We have seen that such freedom is compatible with determinism, but this theory needs a defense against those who would claim it does not constitute real freedom, or at least not any type of freedom we are inclined to care about. To execute this defense, I return to the question I began the chapter with: why are we typically compelled to believe in categorical freedom of choice?[120]

The first reason is that it seems to us while we are deliberating that we are free. In particular, I outlined four phenomenological pieces of data that support the appearance of freedom: 1) An agent's choice appears to her to be the proximate cause of her action. 2) An agent feels that unless she actively engages in a certain sort of activity, no decision will be made. 3) An agent is aware that she will be responsible for the choice she makes, because it seems to her that the outcome of her deliberative process is the only thing that will determine what she will do next. 4) An agent feels, at least in most instances of deliberation, that she is in a position to choose among a variety of genuinely available

[120] I am assuming for the sake of argument that categorical freedom of choice is the most robust and appealing sense of freedom on the market, and so if intelligent deliberation can offer everything that categorical freedom of choice can, intelligent deliberation has nothing more to prove. However, it is worth noting, as many others have, that categorical freedom of choice might itself be an incoherent notion, in which case, so the much better for intelligent deliberation.

options, and that there are no facts, internal or external, forcing her to choose one option over another.

All four of these pieces of introspective data are not only compatible with, but also predicted by prospectivism's theory of deliberation. Imagine the phenomenology of an agent behaving in accordance with prospectivism. To begin with, the majority of her life will not be spent deliberating. In some instances, however, the environment will prevent her from doing what comes naturally, or she will find that she is naturally inclined to behave in two or more conflicting manners. When this happens, she will be thrown into a state of disruption, and in order to return to a state of natural activity, she will need to imagine, or envision, various possible courses of action. In other words, because the need for deliberation was generated by a conflict between her impulses and habits, the only true resolution[121] to the conflict will be the result of her own mental work (this explains 3 above). It will feel to her that this process requires effort, because it requires her not to simply react in a habitual way to familiar stimuli and impulses, but rather to exert attention to the project of envisioning novel ways of acting (this explains 2 above).

To achieve this imaginative task she will rely on her habits of problem solving to identify and assess salient features of the impasse, and her understanding and observation of past sequences of events to predict what consequences would result from taking one course of action or another. Eventually, one of these imagined courses of action will satisfy, in virtue of its representational content, the conditions for settled prospection.

[121] If an agent is deliberating about whether to go to a movie or to a play and she receives word that the play has been cancelled, the conflict disappears, but is not truly resolved. Consider: if Palestine and Israel and all the people therein were to suddenly disappear, there would no longer be a conflict between Israel and Palestine, but it would be strange to say the conflict had been resolved as opposed to simply removed.

From the first-person perspective, the event of this object of thought satisfying these conditions and resolving the deliberative impasse will be experienced as decision, and will be followed by the motion of her body in accordance with that decision. In this way, her choice (i.e. her settling upon an imagined course of action) will appear to her as the proximate cause of her action (explaining 1 above).[122]

Whether prospectivism is compatible with the fourth consideration above requires closer examination. As for the appearance of the lack of internal constraints upon freedom, there are two potential points of conflict with prospectivism. The first would arise if this introspective observation of freedom amounts to appearing to have control over the generation of her thoughts, and the second would arise if this introspective observation amounts to appearing to have control over her motivational reaction to her thoughts.

Imagine an agent who is trying to decide what to do in reaction to the news that a class she was planning to take has been cancelled. Suppose she decides to look through the course catalog and upon doing so, finds a different class, say, World Religions, that she is even more excited about taking and that will fulfill the same requirement. Her deliberative quandary is over: she knows how to proceed. But it is a fair question to ask: to what extent did she have control over her decision? The emergence of the idea of taking World Religions was not in her control, rather it was sparked by a feature of the environment, namely some words on the page of a course catalog. Similarly, the

[122] This is, of course, a description of a successful deliberative agent. Similar stories could be told about how it feels when efforts to settle on a choice are thwarted for one reason or another, or what it is like when an agent manages to proceed with action without reaching complete resolution.

excitement she felt in response to the idea of taking World Religions was not within her control, but rather dictated by motivational dispositions beyond her control.

In this case, I agree with Dewey's assessment of the amount of control we have over our thoughts—to a large extent, the thoughts that spring to our minds when we deliberate are not the product of our conscious choice. He writes,

> Thoughts sprout and vegetate; ideas proliferate. They come from deep unconscious sources…Some suggestion surges from the unknown. Our active body of habits appropriates it. The suggestion then becomes an assertion. It no longer merely comes to us. It is accepted and uttered by us. We act upon it and thereby assume, by implication, its consequences. The stuff of belief and proposition is not originated by us.[123]

Our ideas are not by and large consciously and deliberately generated, but rather, consciously and deliberately appropriated from suggestions that "surge from the unknown." (In principle, the "unknown" here might in fact be an ingenious physiologist.) If I am right about diagnostic reason, one of its chief gifts is that it directs our attention in a non-random fashion toward those propositions and possibilities that are promising. But since diagnostic reason is called for precisely because our antecedent motivations have ceased to effectively guide behavior, we should expect the deliberator to feel as if no one antecedent motivational state will alone determine what she will do.

Thus, in opposition to those who have claimed that the phenomenology of deliberation "certainly points"[124] to categorical freedom of choice or free will, I maintain that nothing in my phenomenology suggests to me that I have categorical control over my thoughts or my imagination during deliberation. Cases of addiction serve to highlight this fact, because they involve unwanted desires. But notice, there is nothing about the act of accepting my desires that makes it appear to me that I manufactured those desires

[123] Dewey, John. 1922/2002. (314)
[124] Taylor, Richard. 1963/1999. (449)

out of an act of conscious will. Wanted and unwanted desires do not appear to come from separate places. For sure, I am glad to be constituted in such a manner that the thought of torturing children repulses me and the thought of helping others pleases me (i.e. I am glad to have the desire not to torture children and the desire to help others). Further, I am aware that, if given the choice, I would certainly choose to be so constituted. But these facts do not imply that I have ever been given such a choice. And in fact, when it comes to my phenomenological data, it seems to me that, for as long as I can remember, these happy facts about who I am have simply been reliably present. I have not had to choose them, or to work consciously to maintain them at all. Since the fourth piece of introspective data above does not itself imply categorical freedom of choice, then we need not conclude that prospectivism conflicts with the phenomenological data.

As for the appearance of the lack of external constraint, this can be explained in terms of the agent's awareness of the presence of ample freedom of opportunity, as defined in the previous section. Normally, when we find ourselves in a situation of deliberative impasse, we also find ourselves in the presence of the resources sufficient to resolve that impasse. Of course, on occasion, we are faced with resources insufficient for the task of meeting the activation conditions without forcibly ignoring certain values or desires, and when we are, we experience this as being <u>forced</u> to act against our wills. This explains the feeling of coercion that accompanies being robbed at gun point, and also the feeling of being forced that no doubt accompanies an agent's decision to make a dangerous jump from a tall, burning building.

In short, I contend that the phenomenology associated with deliberation is relatively unspecific with regards to its metaphysical suggestions and therefore is capable of supporting numerous theories of agency and deliberation. This fact allows many to confuse their beliefs about what freedom of choice involves with what is actually suggested by the phenomenological data associated with deliberative experience. Any time a philosopher claims that it is the determinist's dialectical responsibility to explain away our strong illusions of free will, ask that philosopher to specify what this illusion of free will amounts to and you will find that it amounts to nothing that is in fact incompatible with determinism.

The second reason that we are tempted to believe in free will is our desire to be self-determined and in control. Here again, we need to consider what exactly this desire amounts to. I am glad to be a human because humans have lots of powers and capabilities that other things and creatures do not have. One of the things I can do that a rock cannot is determine for myself where to go and what to do (this is often referred to as 'freedom of action'). One of the things that I can do that a dog cannot (if my intuitions about dogs' mental lives are correct), is to decide what type of life I want to live and what type of individual I want to be. Furthermore, I am glad to be free from serious addictions and compulsions, because one of the things I can do that an addict cannot is act in accordance with what I take to be the most reasonable or beneficial action, all things considered, rather than acting primarily in accordance with a desire which I do not want, or do not identify with, or do not consider important. I claim that my abilities to do these things constitute my capacity for self-determination.

In this way, the degrees of difference between my capacity for self-determination and that of a rock, a dog, or an addict are directly equivalent to the degrees of difference between our respective capacities for intelligent deliberation. There is much more to be said about exactly what is involved in these differences, i.e. about what is involved in being a person as opposed to some other type of creature, and many philosophers have offered theories of personhood that could be used to supplement prospectivism.[125] But for now, I hope merely to demonstrate that while intelligent deliberation cannot satisfy an agent's desire for categorical freedom, it can satisfy the desire to be self-determined in a way that rocks, dogs, and addicts are not. And in fact, when I consider my own desires about my deliberative self, I must admit that I cannot quite understand why having categorical freedom would be any more appealing to someone than having deliberative freedom. Hence, deliberative freedom is capable of supporting a limited, yet satisfactory, type of self-determination.

Finally, we must ask whether prospectivism is enough to serve as a backbone to our understanding of moral blame and moral praise, and to justify our practices of holding people morally responsible for their actions. Surely, if what we seek is to be warranted in believing that humans are, in an absolute and metaphysical sense, the "sole authors"[126] of their own actions, then prospectivism will not suffice. But if what we seek is something else, say to avoid concluding that, all along, we have been behaving in a nonsensical or cruel fashion, then prospectivism is satisfactory.

---

[125] Particularly, I have in mind much of David Velleman's work on agency, deliberation, and will, as well as some aspects of Harry Frankfurt's work.

[126] 'Sole authorship' as a requirement for moral responsibility is C.A. Campbell's concern in his essay, "Has the Self 'Free Will'?" (Campbell, C.A. 1967/1999.)

The suspicion that moral responsibility cannot be justified in the absence of categorical freedom of choice stems from the belief that moral responsibility is incompatible with determinism. The argument goes something like this:

P1) An agent whose is pushed to the floor by an earthquake is not responsible, morally or otherwise, for her falling to the floor. (True by strong intuition.)

P2) The reason she is not responsible is because her behavior is determined entirely by forces beyond her control. (True by strong intuition.)

P3) It is cruel and/or nonsensical to punish or blame an agent for her behavior if that behavior is determined by forces beyond her control. (True by intuition and appeal to the case in P1.)

P4) If determinism is true, then in all instances an agent's behavior is determined entirely by forces beyond her control, namely, the laws of physics and the initial conditions of the universe. (True by definition.)

P5) Determinism is true. (True by supposition.)

C1) In all instances, an agent's behavior is determined by forces beyond her control.

C2) In all instances an agent is not responsible for her behavior.

C3) In all instances it is cruel and/or nonsensical to punish or blame an agent.

Philosophers who want to maintain both that determinism is true and that our moral practices are justified can attack this argument in different ways. For instance, Harry Frankfurt argues against the second premise, insisting instead that an agent's moral responsibility for her behavior has to do with whether she endorses (approves of or identifies with), on a second-order level, the desires and goals that move her to act.[127]

Prospectivism instead attacks the third premise, maintaining that we can make sense out of pronouncements of moral blame and praise even if an agent's actions were determined entirely by the laws of physics, and that in fact, such pronouncements are

[127] Frankfurt, Harry. 1988. See in particular, "Freedom of the Will and the Concept of a Person." (11-25)

often acts of kindness. To understand just the sense in which prospectivism thinks moral praise and blame are warranted, it helps to start with the following admonition from Dewey:

> To content ourselves with pronouncing judgments of merit and demerit without reference to the fact that our judgments are themselves facts which have consequences and that their value depends upon their consequences, is complacently to dodge the moral issue, perhaps even to indulge ourselves in pleasurable passion just as the person we condemn once indulged himself. The moral problem is that of modifying the factors which now influence future results. To change the working character or will of another we have to alter objective conditions which enter into his habits. Our own schemes of judgment, of assigning blame and praise, of awarding punishment and honor, are part of these conditions.[128]

In essence, Dewey holds that acts of moral blame and praise derive their meaning in the same way any course of action derives its meaning—by our taking the course of action into consideration in light of all of its consequences. For Dewey, a prerequisite for morality is the recognition that our reactions to others are themselves part of the social environment; they are among the many factors that shape human habits and contribute to our understanding of ourselves.

Consider, if human minds were as intractable as thunderclouds, morally blaming an arsonist for starting a fire would make as much sense as morally blaming a lightning bolt for the same infraction. But human minds are much more tractable than thunderclouds. For us, the reception of positive or negative judgment from members of our community means something. Morally praising or blaming an agent gives her a reason and an opportunity to reinvent her habits of deliberation to better harmonize with the environments she regularly encounters (or, alternatively, to alter her environment as she sees fit).

---

[128] Dewey, John. 1922/2002. (19) Emphasis his.

For Dewey, determinism does not strip our moral vocabulary of its meaning, nor does it require us to give up our social practices of praising those who act in accordance with our deepest moral principles and blaming those who fail to do so. Rather, it implores us to embrace the fact that our moral attitudes and actions themselves become part of a social environment that, in turn, will provoke our impulses, influence our habits, and ultimately determine how we all behave.[129] Dewey writes:

> We are disapproved, and this disapproval is not an inner state of mind but a most definite act. Others say to us by their deeds "we do not care a fig whether you did this deliberately or not. We intend that you shall deliberate before you do it again, and that if possible your deliberation shall prevent a repetition of this act we object to." The reference in blame and every unfavorable judgment is prospective, not retrospective. Theories about responsibility may become confused, but in practice no one is stupid enough to try to change the past. Approbation and disapprobation are ways of influencing the formation of habits and aims; that is, of influencing future acts. [130]

This passage, taken with the other, expresses the denial of the third premise above. It is not nonsensical to change our environment in a way that better conforms to our ideals. Therefore, if morally praising and blaming agents is a means to changing others in a way that better conforms to our ideals, then these practices are not nonsensical, even if the behavior of the person we are praising or blaming was determined by the laws of physics.

However, this consequentialist approach to explaining the sense and significance of moral practices has significant drawbacks. P.F. Strawson, in his article "Freedom and Resentment" offers an explanation for why some philosophers—he calls them the pessimists—refuse to accept this account. He writes that when moral practices are justified in terms of behavior modification they are

---

[129] Ayer seems to be in agreement with Dewey on this point when he writes "[Even if determinism is true it does not] entail that my actions make no difference to the future: for they are causes as well as effects; so that if they were different their consequences would be different also." (413)

[130] Dewey, John. 1922/2002. (316)

> …represented solely as instruments of policy, as methods of individual treatment and social control. The pessimist recoils from this picture…He is apt to say, among much else, that the humanity of the offender himself is offended by this picture of his condemnation and punishment…[The picture] is painted in a style appropriate to a situation envisaged as wholly dominated by objectivity of attitude.[131]

The general idea here is that treating an agent as merely a system-to-be-modified in accordance with our own wishes is equivalent to treating her as if she is not a member of the moral community. But this behavior—treating members of the moral community as if they are proper objects of our control—is considered by most, including me, a moral offense. And, if the explanation given to justify moral practices is itself morally offensive, it is inadequate. Why should we care that our moral practices are reasonable and meaningful if they are also immoral? Prospectivism's idea of agents as intelligent deliberators can appease the pessimists.

Holding another agent responsible for her actions, on the prospectivist picture, is not a matter of controlling another agent, but rather, a matter of staking out the boundaries of one's normative community. In particular, the verbal or visual expression of praise and blame is a matter of communicating to the other that they have failed to take as authoritative the values that would qualify them as a full member of your normative community.

This can happen on both the trivial and more significant levels. When one housemate blames another for failing to wash her dishes in a timely manner, she communicates the expectation that everybody in the house is to pull their own weight. She is, in essence, saying that, in order to be a full-fledged member of the house, her housemate must adopt and respect that value of a decent shared living space. On a more

[131] Strawson, P.F. 1962/1976. (206)

serious level, when one citizen accuses another of racism, she communicates her expectation that discriminatory attitudes have no place in her society.

The imperatives we express via praise and blame are not, therefore, categorical. They are hypothetical, of the form: in order to be part of sub-community c, you must adopt values v. Expressing such an imperative does not imply the disrespect of the agent toward whom it is aimed, nor an effort to manipulate that agent. Rather, the blame is informational, informing the agent of stakes of which she may have been previously unaware, so that she can, if those stakes are important to her, figure them into her future deliberations. Those who praise and blame us do not manipulate us. Rather, they set us a problem, and thereby give us a choice: we can accept their value propositions as our own and remain in the community, we can reject both the community and the value propositions, or we can try through dialog to convince them that the value expectations they hold are unreasonable or inappropriate.

Our attitudes and expressions of praise and blame thus perform a critical social function. They are part of mapping the terrain of values against which future decisions must be made. Dewey thus gets the essential point correct. Blaming and praising are essentially forward-directed. But they are not forward-directed in the sense that we are trying to manipulate others to our ends. Rather, they express our expectation that another agent is capable of grasping the communal stakes of her decision. This is quite the opposite of what the pessimist feared. Such an expression is an expression of respect, not disrespect. We are not attempting to manipulate agents to our ends. We are attempting to inform them of our expectations.[132]

[132] A similar story can be told about expressions of approval and disapproval.

Thus we can make sense of our moral practices, and be assured that they are not cruel and disrespectful, so long as they effectively embody the aim of informing others of the communal stakes of their decisions. Notice, condemning someone for falling down in an earthquake does not fit this description, and this is what makes punishing someone for such behavior nonsensical and cruel. But punishing people for stealing; holding them morally responsible for breaking promises; praising them for sharing their good fortunes with others; blaming them for failing to react more quickly in a disaster—all of these moral practices do fit the description. These practices are justified and moral.

If praise and blame, and approval and disapproval, are expressions of our own value expectations (i.e., our expectations regarding what will be valued by members of our moral community), then such behavior plays an important reciprocal educational function. When we blame an individual for an action and demand reparations, we not only put at risk their membership in our moral community, but ours in theirs. Since the cost of disintegrating one's moral community is high, trying to jointly reconcile value discrepancies is, by default, the preferable strategy. This will involve, first, diagnosing the various problems that have led us to accept the values we have come to accept and behave in the ways we have come to behave, and second, searching for a shared understanding that will allow us to avoid the positing of irreconcilable value differences. In this way, the practice of holding others morally responsible leads inevitably to a public mapping of the normative territory we all share. This is why the practice is justifiable, even in a deterministic world. Not only can prospectivism's conception of intelligent deliberation give us what we might have once wanted from categorical freedom of choice, it provides a vindicating explanation of our moral practices.

Bibliography

Aarons, Leroy. 1995. Prayers for Bobby: A Mother's Coming to Terms with the Suicide of Her Gay Son. New York: HarperOne.

Anderson, Elizabeth. 1993. Value in Ethics and Economics. Cambridge, MA: Harvard University Press.

Arpaly, Nomy. 2003. Unprincipled Virtue. New York: Oxford University Press.

2006. Merit, Meaning, and Human Bondage. Princeton, NJ: Princeton University Press.

Ayer, A.J. 1969/1999. "Freedom and Necessity" Originally published in Ayer, A.J. Philosophical Essays. Reprinted in Joel Feinberg and Russ Shafer-Landau, Eds. Reason and Responsibility,10th Ed. New York: Wadsworth Publishing Company. (408-413)

Bargh, J. A., and Chartrand, T. L. 1999. "The Unbearable Automaticity of Being" in American Psychologist. Volume 54, Issue 7. (462-479)

Buss, Sarah and Overton, Lee, eds. 2002. Contours of Agency: Essays on Themes from Harry Frankfurt. Cambridge, MA: The MIT Press.

Campbell, C.A. 1967/1999. "Has the Self 'Free Will'?" Excerpted from On Selfhood and Godhood. Reprinted in Perry, John and Bratman, Michael, Eds., Introduction to Philosophy: Classical and Contemporary Readings, 3rd Ed. New York: Oxford University Press. (417-426)

Chi, M. T. H., Feltovich, P., & Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. Cognitive Science, Volume 5. (121-152)

Darwall, Stephen. 2006. The Second-Person Standpoint. Cambridge, MA: Harvard University Press.

Davidson, Donald. 1980/2001. Essays on Actions and Events. New York: Oxford University Press.

Dennett, Daniel C. 1984. Elbow Room: Varieties of Free Will Worth Wanting. Cambridge, MA: The MIT Press.

Dewey, John. 1922/2002. Human Nature and Conduct. Original publication: New York: Henry Holt and Co. Republished: New York: Prometheus Books.

Frankfurt, Harry. 1988. The Importance of What We Care About. New York: Cambridge University Press.

Gibbard, Allan. 1990. Wise Choices, Apt Feelings. Cambridge, MA: Harvard University Press.

2003. Thinking How to Live. Cambridge, MA: Harvard University Press.

Gilbert, Daniel. 2006. Stumbling on Happiness. New York: Alfred A. Knopf.

Gladwell, Malcolm. 2005. Blink. New York: Little, Brown, and Co.

Helm, Bennett. 2001. Emotional Reason. Cambridge: Cambridge University Press.

Hieronymi, Pamela. 2005. "The Wrong Kinds of Reasons" in The Journal of Philosophy, CII, no. 9. (437-457)

Houser, Marc D. 2006. Moral Minds. New York: HarperCollins Publishers.

Korsgaard, Christine. 1996. The Sources of Normativity. Cambridge: Cambridge University Press.

2009. Self Constitution: Agency, Identity and Integrity. New York: Oxford University Press.

Lehrer, Jonah. 2009. How We Decide. New York: Houghton Mifflin Harcourt.

Millgram, Elijah. 1997. Practical Induction. Cambridge, MA: Harvard University Press.

2001. Editor. Varieties of Practical Reason. Cambridge, MA: The MIT Press.

Richardson, Henry. 1997. Practical Reasoning About Final Ends. New York: Cambridge University Press.

Schroeder, Timothy. 2004. Three Faces of Desire. New York: Oxford University Press.

Smith, Michael. 1994. The Moral Problem. Malden: Blackwell Publishing.

Stace, Walter T. 1952/1999. "The Problem of Free Will," excerpted from Stace, Walter T. Religion and the Modern Mind. Reprinted in Joel Feinberg and Russ Shafer-Landau, Eds. Reason and Responsibility,10th Ed. New York: Wadsworth Publishing Company. (413-418)

Strawson, P.F. 1962/1976. "Freedom and Resentment," in Proceedings of the British Academy. Reprint edition, Nendeln/Liechtenstein: Kraus-Thomson Organization Limited. (187-211)

Taylor, Charles. 1985. Human Agency and Language. Cambridge: Cambridge University Press.

Taylor, Richard. 1963/1999. “Freedom and Determinism” in John Perry and Michael Bratman, Eds., Introduction to Philosophy: Classical and Contemporary Readings, 3rd Ed. New York: Oxford University Press.

Thomason, Richmond. 1987. “The multiplicity of belief and desire" in Reasoning About Actions and Plans, M. George and A. Lansky, eds. Los Altos: Morgan Kaufmann.

Velleman, J. David. 2000. The Possibility of Practical Reason. New York: Oxford University Press.

2008. "A Theory of Value" in Ethics, Volume 118, number 3. (410-36)

Vogler, Candace. 2002. Reasonably Vicious. Cambridge, MA: Harvard University Press.

Wallace, R. Jay. 2006. Normativity and the Will. New York: Oxford University Press.

2008. "Practical Reason" in The Stanford Encyclopedia of Philosophy (Summer 2009 Edition), Edward N. Zalta (ed.)
URL = <http://plato.stanford.edu/archives/sum2009/entries/practical-reason/>.

Wegner, Daniel. 2002. The Illusion of a Conscious Will. Cambridge, MA: The MIT Press.

Williams, Bernard. 1979/2001. Originally published in Ross Harrison, ed., Rational Action. Cambridge University Press. Reprinted in Elijah Millgram, ed., Varieties of Practical Reason. Cambridge, MA: The MIT Press.

1996. “History, Morality, and the Test of Reflection” in Korsgaard, Christine. The Sources of Normativity. Cambridge: Cambridge University Press.