6
The Shape of Agency

6.1 Introduction

I have thus far discussed various components of agency: control, non-deviant causation, intentional action. In this chapter I step back to consider the agent in broader context. The discussion helps motivate what I wish to say in the two chapters that follow this one. Those chapters regard modes of agentive excellence: skill and knowledgeable action. Those modes are easier to understand as excellences once we have a deeper understanding of certain basic features of agency: of what being excellent as an agent is excellence in being.

6.2 Standards, Coherence, and Success

Not every system qualifies as an agent. What is special about those that do?

It is an often endorsed, and plausible, idea that to qualify as an agent a system should conform to certain normative standards. Often this is put in terms of rationality. So, for example, Donald Davidson claims that "An agent cannot fail to comport most of the time with the basic norms of rationality" (Davidson 2004: 196–7). And Christian List and Phillip Pettit claim the following:

> The very idea of an agent is associated with some standards of performance or functioning, which we call "standards of rationality." These must be satisfied at some minimal level if a system is to count as an agent at all.
>
> (List and Pettit 2011: 24)

This sounds plausible. But we might want a more encompassing notion than that given by rationality. Consider a very simple system—the paramecium, a single-celled eukaryote. It has some internal structure. It has some causal powers. It moves through certain liquids by the coordinated beating of its cilia. It can navigate around obstacles or escape certain substances by

The Shape of Agency: Control, Action, Skill, Knowledge. Joshua Shepherd, Oxford University Press (2021).
 DOI: 10.1093/oso/9780198866411.003.0006

way of an "avoiding reaction"—a reversal of its direction of movement, and a slight changing of course in some random direction (Kung and Saimi 1982). This is not a very efficient way of navigating, but the thing is stuck with very short cilia. In any case it is also capable of reproducing, and its methods appear good enough for evolution to keep it employed—many a paramecia has survived long enough to reproduce.

Most philosophers would not want to apply a notion of rationality to the movements of the paramecium. But we might still apply standards of success to its behavior.

In virtue of what? In virtue of the needs or functions we might, as biological theorists constructing a good biological theory or a fruitful model of the thing, impute to it. The constitution of the system—its causal powers, its conditions for continued existence or for reproduction or for functional contribution in some larger system—set in a context of some set of chosen or typical circumstances, sets up a space of standards for that system's behavior. In the case of the paramecium, we might want to model its navigation capacities, and we might impute a goal to avoid some obstacle or some substance as a way of doing so. Whether the goal is a legitimate standard of success for the paramecium may depend upon the quality of our model—whether navigating the obstacle contributes to survival or reproduction rates.

A system might meet imputed behavioral standards in a way that does not even trend in the direction of agency. Consider a system only capable of moving in one direction along a flat surface. We might impute a function to this system, based on its survival needs: the system needs to find and fall into small gaps in a surface on which it moves. If it fits into the gap, it wins. The system does so in the only way it can, by moving blindly in one direction along the wall. There aren't so many small gaps in the wall, but every once in a long while—make it as unlikely as you like—it comes across one. It wins. This may be enough for the system. Such a system does not trend in the direction of agency.

The paramecium does, however. Indeed, Tyler Burge argues that "primitive agency" extends down to the level, at least, of single-celled eukaryotes. Burge points to the orientation behavior of such organisms:

> Taxes are directional movements with respect to stimulations in the environment. They require sensory capacities that are directional. Usually determining direction depends on there being two or more locations of sensory receptors on the body of the organism. Directional movement is

> usually achieved by some mechanism in the animal for simultaneous differentiation of intensities of stimulus registration in different bodily sensors. For example, the animal might turn toward or away from the direction perpendicular to the side of its body that receives the most intense stimulus registration. (Burge 2009: 258)

Burge judges that coordinated, functioning orientation behavior of simple organisms—e.g., "The paramecium's swimming through the beating of its cilia, in a coordinated way, and perhaps its initial reversal of direction" (259)—qualify them as agents. As Burge writes, "Such organisms are capable of steering toward or away from a stimulus source, subsequent to internal differentiations between stimulus intensities in different areas of the body" (258). The movement toward a stimulus is caused in a different way than the movement away from a stimulus, and the difference makes sense in light of the system's own activity—the transitions between states of the system that are differentially sensitive to stimulus source and intensity. That is, the system's behavior is not only reliably produced, it is coherently produced given the circumstances. And it permits something like success. The system's behavior is related to imputable goals regarding their needs (for safety, for finding energy sources, or whatever) with respect to their environment. In their typical behavioral circumstances, this orientation behavior reliably leads to successful (enough) approximation of these goals.

Many will disagree with Burge that we find agency at this level.[1] After all, reproduction is no less complicated and important a process for the

[1] Consider, however, John Hyman's impressively sparse view on (non-intentional) action and agency. For Hyman, Agency is present whenever there is action, and action is present whenever a substance causes a change. Thus, we see action "whenever we see someone walking or speaking…when we feel the sun warming our skin or ice cooling our tongue" (Hyman 2015: 29) One might worry, of course, that Hyman is here changing the subject, or considering a different one. But Hyman does go on to define notions of activity and passivity in complex agents. Activity in the life of complex agents is understood as action that proceeds from the functioning integration of the system's parts:

> The proportion of human acts that stem from an intentional movement is neither here nor there, as far as defining the distinction between activity and passivity is concerned. For the agency of complex substances with functionally differentiated parts always depends on the integrated operation of these parts, rather than on the operation of a specific part or faculty—e.g. the amygdala or the will. This applies to human beings in the same way as it applies to other animals and colonies of animals, and all other complex agents, including institutions and machines. Hence if some basic human activity that involves the integrated operation of cognitive and motor systems, such as feeding or copulating, were only conscious and controllable to the extent that breathing is, a smaller proportion of human acts would stem from an intentional movement than is actually the case, but activity and passivity in human life would be distinguished in exactly the same way. (Hyman 2015: 52)

paramecium than is locomotion. But it is less intuitive to think of a paramecium's asexual reproduction, by a process of binary fission, as an example of primitive agency. That's just the mechanics of life. And if so, why not the beating of the cilia, or the avoiding reaction (which, by the way, often occurs spontaneously)?

Whatever we think about the agency of a paramecium, Burge is right to emphasize continuity between this level and others. At this low level we find key ingredients of agency. Behavioral standards must be imputable to the system. Behavior must be coherent in light of the relevant behavioral standards. And behavior must be reliable in meeting or approximating these standards—the system must succeed, to some degree.

One could try to forge imputability, coherence, and reliability into something like necessary conditions for agency. I decline to do so here, in part because doing so would require a lot of work, in part because I am not sure of the prospects for success, and in part because I do not need to do so. These are, I think, key elements of at least many agents. And these features vary in sophistication, such that more sophisticated agents correlate with increases in the standards imputable, changes in the nature of the standards, differences in the methods by which coherence at meeting the standards is reached, and more sophisticated layering in the mechanisms that help secure reliability in behavior.

6.3 Psychological Agency

A level up from systems like the paramecium, behavioral standards that apply to the system can still be drawn from the system's needs or functions. But at this level the system begins to set certain standards for itself. For at this level the system has the capacity to represent objects, and goals for action regarding these objects—to token psychological states that link it and its behavior to the world in reliable ways. Burge here invokes the notion of a perspective:

> When perception sets an object for animal action, agency reaches a new level of sophistication. The action is suited to a goal that the animal itself

This is an interesting view. But I worry it explains too little. Functional integration is obviously important for any system to qualify as a system. And the notion of a function may give us some indication of what is really important here: the application of behavioral standards to the system. That is the notion that seems to be missing, or too nascent, in Hyman's view.

> perceptually represents. If an animal can perceive, it has some perspective on its objectives. (Burge 2009: 267)

One might think that this is the level at which agency truly emerges. This is what Sean Thomas Foran (1997) argues. According to Foran, an animal moves itself, as opposed to being passively moved as a rock is, when the animal's movements are shaped with respect to objects of that animal's perception. Foran's notion of movement being shaped seems similar to the notion I offered just above, of a system's behavior being coherently produced:

> "Movement shaped with respect to an object of perception" does not simply mean "movement caused by perception." Movement can be caused, in some quite general sense of "caused," by perception without being shaped with respect to the object of that perception. Consider this example. Suppose that when a certain kind of quadruped animal sees one of its natural predators, it immediately lowers itself to the ground and remains still. Perceiving the predator causes the animal to lower itself, but the movement that is caused is not shaped with respect to the predator. The movement is still shaped with respect to something the animal perceives, the ground, but its perception of the ground is not what led it to lower itself: this episode of movement was caused by perceiving the predator.
>
> (Foran 1997: 41–2)

At this level, perhaps, it becomes appropriate to think of coherent production of behavior in terms of practical rationality. When a system can represent behavioral targets, and can implement plans for behavior that approximate standards of success regarding these targets, that system's behavior might well be considered practically rational. And some of that system's behavior might be considered intentional action.

We are still, however, at a level of relative simplicity. At this level it is important that the system be embedded in circumstances in the right ways. For, while the system may be able to represent targets for behavior and deploy plans to hit these targets, the behavioral profiles deployed in following the plan may be inflexible. And inflexible behavioral profiles contain a flaw regarding the meeting of certain behavioral standards.

Distinguish between success according to the standard a system's particular goal or plan sets, and success according to the standards that apply to that system more broadly. If the system is at all complex, then the standards that apply to it will be broader than the standards a particular goal or

plan sets. It will have a range of needs, or perform a range of functions.[2] It may even have a range of intentions, which need to be delicately executed in order not to fail with respect to some of them. Inflexible behavioral routines lock the system into one way of behaving, making it difficult for the system to change tack, or to adjust even slightly. As a result any infelicitous circumstances, or any kinks in the plan, may throw the system off course.

Consider the digger wasp, Sphex ichneumoneus. In preparing to lay her eggs, the Sphex displays some extraordinarily intelligent-seeming behavior. It catches and drags a cricket into its burrow, lays eggs in the burrow, closes the burrow, and leaves.

Regarding this behavior, Woolridge (quoted in Dennett 1984: 11) comments (though the actual details regarding Sphex behavior may be more complicated—see Keijzer 2013):

> To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness—until more details are examined. For example, the wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go inside to see that all is well, emerge, and then drag the cricket in. If the cricket is moved a few inches away while the wasp is inside making her preliminary inspection, the wasp, on emerging from the burrow, will bring the cricket back to the threshold, but not inside, and will then repeat the preparatory procedure of entering the burrow to see that everything is alright. (Woolridge 1963: 82)

Apparently, the Sphex will do this repeatedly, no matter how many times one tampers with the cricket. Commenting on the Sphex's strange behavior, Dennett writes: "Lower animals, such as Sphex, are constitutionally oblivious to many of the reasons that concern them" (Dennet 1984: 24) By reasons Dennett is referring to certain courses of action rationalized by the animal's

[2] This is certainly true of human agency. I say this with Neil Sinhababu's partial account of human agency in mind. Sinhababu advances the following claim:

> Humean Self-Constitution: Agents are constituted in part by all of their desires, and aren't constituted by any other motivational states. (Sinhababu 2017: 167)

Sinhababu continues: "Humean Self-Constitution is neutral about which other non-motivational mental states also go into constituting an agent, or whether any do" (167). My suggestion to Sinhababu would be that one cannot capture human agency without capturing the way humans meet the behavioral standards that apply. At least Sinhababuian intentions—which include means-end beliefs fused to desires—would be required, plus relevant behavioral (i.e., control) capacities.

own background needs, drives, and (if such states can be legitimately attributed to the animal) beliefs and desires. One problem with the Sphex's behavior is it appears blind to a wide range of pressing practical reasons, in the sense that the animal can be placed in circumstances that render it systematically poor at achieving its own basic goals.

This may lead one to reject the idea that such animals can be assessed according to standards of practical rationality. Running against this rejection is the fact that all animals, including humans—such is the lesson of many years of research on human reasoning and its pitfalls (see Gilovich, Griffin, and Kahneman 2002)—can be placed in circumstances that render them systematically poor at achieving behavioral goals.

Susan Hurley argues that many non-human animals can be assessed according to rational standards, but only in certain circumstances:

> An intentional agent who lacks context-free conceptual and inferential abilities and does not conceptualize her reasons can still act for reasons that are her own, reasons from her perspective. Her point of view may provide islands of practical rationality rather than a continuous space of reasons. Reasons for action can be context-bound and lack conceptual generality. (Hurley 2003: 231)

Hurley illustrates these claims by contrasting two kinds of experiments involving chimpanzees. In one experiment, a chimpanzee sees that another chimpanzee (the "indicator") has visual access to the location of a reward in a box, and learns to access the reward by following what the indicator indicates. Then the chimpanzee sees that the boxes are switched, unbeknownst to the indicator, who then indicates where the reward is. The chimpanzee should infer that the reward is in the box that the deceived indicator failed to indicate. But the chimpanzee never does. She follows the indicator's advice and receives no reward. It seems as though the chimpanzee has no access to a crucial reason for action, and that she would have access if she were tracking the relation of the mental states of the indicator to the location of the reward. The chimpanzee must be tracking something else.

And yet, in a second experiment, dominant and subordinate chimpanzees are put in the position to access food. The subordinate chimpanzees are very good at noticing when the dominant has or has not seen the food. When the dominant has not seen the food, the subordinates tend to go and safely access the food. In this case it seems as though the chimpanzee has access to a crucial reason for action, and that this access stems from her

tracking the relation of the mental states of the dominant to the location of the food. Hurley comments: "It may be natural for chimps to compete over food; their ability to act rationally in light of the mental states of others may be evolutionarily tuned to competitive practical contexts rather than cooperative ones" (Hurley 2003: 250–1).[3]

Whatever the proper psychological explanation, the case illustrates the possibility of a system that can be assessable by reference to the norms of practical rationality in certain special circumstances but not in others (cf. Morton 2017).

Now, the range of circumstances in which a system can follow or approximate various behavioral standards will probably vary by degree. In biological creatures, increasingly sophisticated psychological structures correlate with a wider range of behavioral success. The insight here was articulated by Paul Grice in a comment on his method of "creature construction"—a method of modelling psychological systems (which he, in reference to Carnap, calls pirots):

> "What are the general principles exemplified, in creature-construction, in progressing from one type of pirot to a higher type? What kinds of steps are being made?" The kinds of step with which I shall deal here are those which culminate in a licence to include, within the specification of the content of the psychological states of certain pirots, a range of expressions which would be inappropriate with respect to lower pirots; such expressions include connectives, quantifiers, temporal modifiers, mood-indicators, modal operators, and (importantly) names of psychological states like 'judge' and 'will'; expressions the availability of which leads to the structural enrichment of specifications of content. In general, these steps will be ones by which items or ideas which have, initially, a legitimate place outside the scope of psychological instantiables (or, if you will, the expressions for which occur legitimately outside the scope of psychological verbs) come to have a legitimate place within the scope of such instantiables: steps by which (one might say) such items or ideas come to be internalized.
>
> (Grice 1974: 41)

As Grice envisions it, higher and higher forms of psychological system evince psychological states with more and more sophisticated structures.

[3] But see Melis et al. 2006.

Some animals appear to follow rules that approximate or embody norms of practical rationality only in some circumstances simply because the representational systems they use to track the world and drive behavior are not able to deploy rules that perfectly mirror practical norms across all circumstances. The honeybee has evolved a richly combinatorial communicative system—the waggle dance—and a good navigational system. The properties of one honeybee's waggle dance will tell other honeybees where to go to find nectar. But consider a series of experiments in which Gould and Gould (1988) (and Tautz et al. (2004)) had honeybees discover nectar in the middle of the lake, which they then reported to their colleagues. Almost as if they didn't believe what they were seeing, the honeybees ignored the waggle dance. One interpretation of this, as Camp (2009) notes, is that the bees put the states <nectar is there> and <there is lake> together into the state <nectar in lake>, which they subsequently rejected. But an alternative interpretation is that the bees failed to make sense of what they saw because of a limit in their representational system. As Camp puts it, "Perhaps their representation nectar there is blocked from interacting with their cognitive map, because the region on the map marked 'lake' can't receive any other markers" (299). If that is right, then the bees have a representational limit that renders them unable to accord with the relevant norm in one circumstance, even though their representational system is overall well-tuned to deliver success.

It is not my aim here to articulate the metaphorical ladder of agency in any fine-grained way, but it is worth mentioning that we seem to find psychological activity in some animals that approximates practical reasoning, without really qualifying as such. There is a level of psychological complexity in between that of rigid behavioral profiles in characteristic circumstances, and full-blown reasoning of the sort humans sometimes use well.

Anthony Dickinson has argued that many animals fail to instantiate imperial cognition, which he understands as taking "the form of propositional-like or explicitly symbolic representations that are deployed in the control of behavior by processes that conform to some normative standard, such as conditional and Bayesian reasoning" (Dickinson 2012: 2733). But Dickinson argues that these animals can approximate such cognition by deploying associative processes constrained by relatively sophisticated processing architectures. In certain circumstances, for example, rats display behavior that is interpretable as reasoning by disjunctive syllogism. If you give a rat an orange-lime drink sweetened with sugar, and then give it an unsweetened lime drink, the rat goes on to prefer an (unsweetened)

orange to an (unsweetened) lime drink: as though the rat reasoned from the thoughts that <the sweetness derives from the orange or the lime>, and <not the lime>, to the conclusion that <the sweetness derives from the orange>. It turns out, however, that the rule the rat is following is not best captured as reasoning by disjunctive syllogism. Rather, the rat depends upon associations between presently perceived items as well as associations between items retrieved by memory (items generated by the present context) (see Dwyer et al. 1998).[4] This allows a layered form of associative learning that, while not as rational across circumstances as disjunctive syllogism would be, is in fact pretty good. As Dickinson explains, "this form of learning greatly enhances the apparent inferential power of the associative system" (2735)—the associative processes are structured in such a way that though they do not mirror the relevant norm confronting the rat, they do "finesse a problem that would seem to require a rational solution" (2736). The rules the rat follows approximate norms of (practical) rationality rather than embodying them.

So we have an idea, now, of systems that at least approximate norms of practical rationality partly in virtue of their representational and psychological capacities. With higher animals, including humans, we come upon greater representational sophistication, and probably upon a new level of agentive sophistication.

6.4 Higher Animal Agency

The kind of psychological agency that many lower animals display is sensitive to applicable behavioral norms in many circumstances, even if the agents themselves do not represent, or intentionally follow, those norms. We expect better of some higher animals. Some higher animals are capable of deploying reasoning that embodies and mirrors the very norms of practical rationality that apply to the animal in her circumstance.

[4] The way the rat actually works through the problem appears to be this. The rat first associates the things it drinks (sweet orange-and-lime) with each other and the surrounding context. When the rat is then given unsweetened lime drink, sweet-and-orange is also brought to mind as a part of the associative link, which further strengthens the connection with orange and sweetness (in spite of the lime's being unsweetened). So when given the option to drink orange in the future, rats who had been given the lime drink would be more likely to drink it than rats who had not, simply in virtue of the strengthened connection between orange and sweetness.

Consider, for example, a system that forms rational plan-states in part by way of genuine practical inferences. These are genuinely inferential transitions from one set of mental states to another, where the sets of mental states involve practical premises and conclusions—premises about what might be done, and conclusions about what to do. On Quilty-Dunn and Mandelbaum's (2018) view of inferential transitions, inferential transitions between mental states are rule-based in that they "constitutively obey some logic" (537), and the explanation for their doing so is that the transitions are sensitive to—occur because of—semantic relationships between the discursive content of the mental states involved.

> If transitions between thoughts are sensitive to constituent structure, those transitions must obey some logic. This is true because a logical rule just is a kind of rule which is sensitive to constituent structures. For instance, suppose that a rule of mental logic is the following: If X is an AN, then X is an N. Suppose, further, that you token the thought BERTHA IS A BROWN COW. You will then, ceteris paribus, token the thought BERTHA IS A COW. The transition is logical because it occurs in virtue of the fact that the constituent structure of the input representation satisfies the antecedent of the rule, and the output is generated because its constituent structure satisfies the consequent of the rule. Transitions between discursive representations that are triggered because their constituent structures instantiate some rule of mental logic thus suffice to make those transitions rule-based and logic-obeying.
>
> (Quilty-Dunn and Mandelbaum 2018:538)

Practical inferential transitions that obey a basic logic are one example of a way that a system could follow norms of practical rationality. A system might have a goal to eat ripe fruit that, in conditions of foraging, is represented as a conditional: <if ripe, then eat>. Perception of ripe fruit might, in conjunction with this goal, lead to a practical inference: <eat (that) fruit>. A system might possess a goal to break a window, but represent the window as too thick to break by hand. This might lead to a transition from the goal to break a window to a sub-goal to scan of the room, by way of a reliable rule connecting the absence of means with a search process. The sub-behavior of scanning might thus be practically rational given the reliability of the rule followed. And this scan might lead to the perception of a hammer as salient to breaking windows, which might lead to a transition from the goal to break a window and the perception of a hammer that can break windows, to

the goal to break a window with that hammer. That kind of transition might follow a rule connecting means to ends, and it might do so because of the constituent structure of the relevant representations. That is, the system might follow the rule in this case because its representation of the hammer—as, e.g., hard, heavy, swingable, or whatever—has a structure that enables it to interact with the represented goal to break a window, and probably with background knowledge of physics, and thus to generate a rule-following transition to the goal to break a window with that hammer. Following such rules in these ways leads the system to accord with norms of practical rationality in the relevant circumstances, and following these rules across a broader set of circumstances is likely to lead to accordance with relevant norms more broadly.

Of course this discussion of following rules like modus ponens assumes that the system in question represents the world and its own abilities well enough to connect antecedent and consequent in appropriate ways. (Following the rule "if ripe, turn around three times and clap your hands" is not likely to lead to accordance with the norms of practical rationality in many circumstances.) It also assumes the system in question possesses a representational system (or systems) that operate in part via logical rules sensitive to the structure of its representations. As such, the deployment of such rules requires a fairly psychologically sophisticated system.

Importantly, though, the deployment of such rules does not require meta-cognition—the system need not represent its own (first-order) representations. It need not possess any explicit understanding of the norms it follows. As Peacocke has noted, much of our reasoning makes no reference to our own attitudes. The primary concern is relationships of reason that obtain in the world. Peacocke distinguishes between first-tier thought, which is about the world but involves no consideration of relations of reason, and second-tier thought, which "involves thought about relations of support, evidence or consequence between contents" (1996: 130) and which leads to rational mental state revision, without involving higher-order thought of any sort:

> It seems to me a four-year-old child could engage in the following simple piece of second-tier thought. First he thinks that a particular toy is in the cupboard, because that is where it is normally kept; then he remembers that his aunt is staying with them, and that she puts toys in the wrong place. He realizes that the cupboard's being the normal storage place does

> not mean the toy is there now, given the presence of Auntie, and so he no longer believes that the toy is in the cupboard. (Peacocke 1996: 130)

With second-tier thought, however, an additional, potentially very useful, layer of sophistication is acquired.[5] Consider Burge's distinction between critical reasoning and reasoning:

> A non-critical reasoner reasons blind, without appreciating reasons as reasons. Animals and small children reason in this way…Not all reasoning by critical reasoners is critical. Much of our reasoning is blind, poorly accessible, and unaware. We change attitudes in rational ways without having much sense of what we are doing. Often, we are poor at saying what our reasoning is. Still, the ability to take rational control of one's reasoning is crucial in many enterprises – in giving a proof, in thinking through a plan, in constructing a theory, in engaging in debate. For reasoning to be critical, it must sometimes involve actual awareness and review of reasons. (Burge 2013: 74)

At this stage—a stage of adult human sophistication that can involve reflection on our reasons as reasons, and that can involve considerations of relations of reason between our various psychological states, we find a level of agency that Michael Bratman has developed in much detail.[6] We find planning agents (see Bratman 1999, 2007). Bratman is not claiming that only planning agents possess plans. Rather, planning agents are agents with sophisticated psychological architectures, and agents the elucidation of

[5] Philip Pettit has designed an account of reasoning that relies on second-tier thought. According to Pettit, reasoning requires a capacity for meta-propositional attitudes—attitudes towards complex propositions "in which propositions may themselves figure as objects of which properties and relations are predicated" (2007: 498). Such attitudes allow agents to assess their first-order attitudes for features like truth, consistency, and entailment. Thus:

> To be able to reason, under the model I shall adopt here, is to be able to conduct an intentional activity that is designed—and perhaps explicitly intended—to raise the chance of satisfying [desiderata of rationality]. Specifically, it is to be able to ask oneself questions about certain propositional properties and relations; to be able thereby to let beliefs form on such matters; and to be disposed to adjust rationally to whatever beliefs one forms. (Pettit 2007: 499)

Of course human agents do sometimes follow norms in this way. But I think we do so in less explicit and intentional ways as well.

[6] See also R. Jay Wallace (1999): "A person is guided by their conception of their reasons when that conception is reflected in the content of the intention on which they act; in that case, one will be able to understand what the agent is doing only by grasping what speaks in favor of so acting, from the agent's own point of view" (239).

which requires great attention to the structure and normative underpinnings of planning activity:

> In support of both the cross-temporal and the social organization of our agency, and in ways that are compatible with our cognitive and epistemic limits, we settle on partial and largely future-directed plans. These plans pose problems of means and preliminary steps, filter solutions to those problems, and guide action. As we might say, we are almost always already involved in temporally extended planning agency in which our practical thinking is framed by a background of somewhat settled prior plans.
>
> (Bratman 2018: 202)

As Bratman's work elucidates, planning agents are veritably bathed in applicable practical norms. We spend much of our time working through implications of our commitments, testing them against other possible commitments, wondering whether some other course of action might be better in some way, wondering how the plan will impact others, or whether refinements to the plan might make profit along some unforeseen dimension.

The capacity of planning agents to recognize norms for what they are and to form plans influenced by these norms can give the impression that the rationality of planning agents is not sensitive to context.

But even planning agents—at least imperfect ones like humans—must be imbedded in appropriate contexts. The psychological apparatus of humans, after all, piggy backs on the apparatus of our ancestors. We, too, often rely on processes that push us at best toward approximation of the norms of practical rationality. We lean on heuristics, we are sensitive to nudges, we display systematic patterns of bias. This is a chief lesson of the social psychology and behavioral economics literatures of the past fifty years. The lesson, in part, is that certain types of circumstances shape our reasoning in systematic ways, and that while this reasoning can work well, there also exist circumstances in which the reasoning is quite poor indeed.

Jennifer Morton (2017) has recently offered research on reasoning in resource-poor and in resource-rich contexts as a vivid example. Agents in resource-poor contexts display different patterns of reasoning—different patterns of focus, different weighting of future versus near-term goods, different valuing of the resources at hand. Morton argues that the difference in circumstances changes—appropriately—the structure of an agent's practical reasoning:

> Deliberation is something that we do in the non-ideal and messy contexts in which we live. Theories of rationality that abstract so far from the context of deliberation as to render it invisible run the risk of ignoring or, worse, of distorting the experiences of those who exercise their agency in conditions of considerable disadvantage. This is not to endorse relativism about rationality. According to the account that I propose, there are facts of the matter about how it is better to deliberate in one context as opposed to another. (Morton 2017: 557)

Morton is right. The norms that apply to an agent are a function of the agent's constitution—their causal powers, their psychological faculties—embedded in a certain set of circumstances. Perhaps the agency of some angels or gods can be characterized with respect to the set of all circumstances. But imperfect agents like humans are, as all other animals are, embedded in sets of circumstances smaller than the set of all of them. This has important consequences for how we think of their agency.

6.5 Conclusion

In this chapter I have charted a metaphorical ladder of agency. I began by noting key elements of agency—the imputability of behavioral standards, the capacity to display coherence in meeting these standards, and the capacity to do so reliably. Steps up the ladder are marked by important structural and featural differences. Animals with certain representational capacities (e.g., perception) can set behavioral standards for themselves. Animals with increasingly complicated and sophisticated psychological architectures develop methods for meeting these standards with more flexibility, across greater differences in circumstance. And at a certain point we begin to find animals capable of reasoning, and even capable of monitoring and reflecting upon the quality of their own reasoning. At this level an understanding of agency and agentive excellence requires some understanding of activities of planning.

It is worth here considering an upshot of this discussion for the following item: the place of mental action in an understanding of agency.[7] Not every agent engages in mental actions undergirding practical reasoning. But

[7] In a forthcoming chapter (Shepherd forthcoming-a) I consider the implications of this general picture of agency for the "problem" of disappearing agents. It turns out not to be particularly troublesome.

humans frequently do. Is there anything special about the actional nature of such processes?

I do not think mental action is any more essential to agency than bodily action (see Levy 2018 for an interesting perspective). But I do think mental action is tied to human agency (and to agents with similar psychological structure) in an intimate way. In particular, I think the pervasiveness of mental action in our mental lives is a product of our particular computational, informational, and cognitive architectural limitations, as well as the solutions evolution seems to have bequeathed to us. I cannot argue the point in full here, but it seems to me that much of our mental action—and especially the actions that contribute to processes of practical deliberation—is driven by uncertainty and conflict (Shepherd 2015a). This uncertainty and conflict are related to our sense of the norms of practical rationality—often, in deliberation, we are engaged in a search to uncover what it is best to do, or how best to execute a pre-existing intention, or how best to navigate a conflict between various desires, or obligations, or commitments, or whatever. We deliberate because we are informationally fragmented in certain ways (Egan 2008)—it is a struggle to call to mind the relevant items, and to put them together in rationally satisfying ways.

This is a feature of our situation, but it need not transmit to agency qua agency. To see what I mean by this, consider a being constitutively incapable of uncertainty or conflict: an omniscient, omnipotent, and fully practically rational being. Call it Al. It is certainly conceivable that Al, in virtue of its supreme knowledge, never faces uncertainty. And Al, in virtue of its full practical rationality, never faces conflict (unless it be a conflict in the very norms of practical rationality). In whatever the situation, no matter how complex, Al need not deliberate to discern the best course of action. We might say that no matter the situation, no matter how complicated or fraught with moral gravity, Al simply sees—takes in at a glance—the thing to do. Al always acquires intentions reflective of Al's omniscience and full practical rationality. (In order to take in all the information required to always discern the thing to do "at a glance," Al will need some pretty amazing perceptual sensitivity and some pretty amazing cognitive sophistication. We can assume this is covered by Al's omnipotence.)

It seems to follow that neither the kind of uncertainty and conflict that is our normal situation, nor the actional processes of deliberation and decision via which we attempt to reduce uncertainty and accord with norms of practical rationality, are essential for agency (Arpaly and Schroeder 2012 and Morton 2017 make the same point). Further, it seems to follow that

uncertainty, deliberation and decision are important features of our—that is, human—agency precisely because human agency is far from perfect. We have perceptual, cognitive, and volitional limitations, and it is because of this that uncertainty, deliberation and decision play such a large role in our lives.

Even if these kinds of mental actions are inessential to the nature of agency, for agents like humans, the activity of practical reasoning that is essential to our agency is often conducted via mental actions—intentional mental activities like shifts of attention, inhibition of urges, imagination of possibilities for action or consequences of courses of behavior, comparison of action options, weighing of reasons, and so on. I want to suggest that for human agency, mental action is a rational response to the computational, informational, and architectural limitations we face. Mental action is a kind of rational glue—it is the way that we attempt to discover the norms of practical rationality, and to enforce rational coherence across the large but disjointed set of goals, preferences, and abilities that we tend to possess.

A final point turns us in the direction of the next two chapters. My concern there is modes of agentive excellence. Excellence is a kind of perfection of form. So it helps to understand the form of agency. As I have it, this form is that of a system whose behavior, internal and external, is integrated in a way that enables the meeting of behavioral standards. As we will see, skill is a refinement of this—a meeting and even an exceeding of behavioral standards along ever finer dimensions. And knowledgeable action is an important method for meeting standards in otherwise difficult contexts. Knowledge assists an agent in expanding the range of contexts in which she might succeed.

With apologies for waxing in this direction, the lesson of this chapter combined with the next two may be put thusly. Hamlet was only part right:

> What a piece of work is man, how noble in reason, how infinite in faculty, In form and moving how express and admirable, in action how like an Angel, in apprehension how like a god, the beauty of the world, the paragon of animals. (Shakespeare 1996: 112)

Of course Hamlet, with a penchant for drama, was overselling things. In truth human agents are deeply flawed. Not infinite in faculty. On some days, in action not much like an angel, in apprehension closer to a macaque. And yet, we do, in some contexts and in some domains, manage movement both express and admirable, and we do, in skilled and in knowledgeable action, close the gap between us and the gods.