**[NOTE: These comments were the basis of a zoom conversation at the Eastern APA, January 2021. Josh is putting them up on philpapers (with permission from all involved) in case they are helpful to people interested in the themes of this book.]**

**APA Author meets critics for Joshua Shepherd, *The Shape of Agency: Control, Action, Skill, Knowledge*, OUP 2021.**

**Andrei Buckareff (Marist): Chair**

**Kim Frost (UC-Riverside): First commenter**

**Sarah Paul (NYU-Abu Dhabi): Second commenter**

**Joshua Shepherd (Carleton University, University of Barcelona): Response**

**Kim Frost’s comments.**

*0. Introduction.*

Joshua’s book describes a relatively complete framework for “fruitful theorizing” about agency, centered around accounts of control, non-deviant causation, intentional action, skill, and knowledgeable action. *Causal models* play a crucial role in understanding the accounts. The book is ambitious: each account I mentioned could fund a book by itself. Anyone interested in agency will find much to engage with in this book, and I’m sure that will be proved in our discussion.

I don’t have much time, so unfortunately I have to be picky. First, I’ll argue that *control’s exercise* is not really a necessary condition of intentional action, even if it’s there in most cases. Then I’ll raise some related clarificatory questions about the title and spirit of the book and its *explanatory goals*. At the end, I’ll raise doubts about whether Joshua has solved the problem of *causal deviance*.

*1. Control and Clumsiness.*

Control’s possession and exercise are central to the book. Joshua (64) says that if an agent performs an intentional action, they must approximate a relevant aspect of their plan because they exercise a sufficient degree of control over their behavior.[[1]](#footnote-1) I think this is probably false.

There’s a reading of the claim on which it’s innocuous and true. On this reading “sufficient degree of control” is just a generic causal notion implying that the agent was the primary source of the relevant desired change, and nothing else had a better claim to be the primary source. So in Davidson’s famous case of causal deviance, where desire to let go of a rope so unnerves a mountaineer that her grip loosens and she lets go, the agent’s nerves (or “unnerving processes”) have a better claim to be the source than the agent herself does. You don’t really need the word “control” to express this idea. On the innocuous reading, “sufficient degree of control” means almost, but not quite, nothing.

I think Joshua wants something much stronger than that from “sufficient degree of control”. I think he wants to rule out *lucky, unusual, clumsy, uncoordinated actions* from being genuine intentional actions.

We don’t have to look far for an example of what I mean. Joshua (31) opens the chapter on causal deviance by recounting the last moments of the last basketball game of his high school career. The clock is about to run down. An opponent nervously leaps into the air with his back to the basket, produces an awkward, ugly-looking flip shot, and, assisted by the rim and a wild degree of spin, manages to get the ball in, despite the terrible angle of the throw. He wins the game by a tiny margin.

Sinking the basket is clearly *very uncontrolled*: to the extent that Joshua (31) wants to deny that it was an intentional action. Certainly if there were many do-overs in this kind of situation, the failures would vastly outnumber the successes. From the vivid way it’s recounted, one gets the impression that Joshua wants a do-over. As I read, I could almost hear him yelling “Oh come on!”

Given Joshua’s definitions, this basketballer does not exercise (a sufficient degree of) relevant control. There are two important conditions on Joshua’s account of control’s exercise, which it inherits (46) from his account of non-deviant causation. First the “level of content-approximation must … be something the agent can do repeatedly, in circumstances like this.” (34) So the basketballer does not exercise control (to a sufficient degree), because he can’t perform similarly, sinking baskets, or getting close, in many do overs. Second “the causal pathways operative in the actual case must be such that taking those pathways leads to something close to the behavior in question, in circumstances like this.” (34) This is, roughly, a reliability condition on the causal pathways *operative* in the given case. We would probably need to know more about the basketballer’s nervousness, and how those pathways enable or mess with (near) successes, to know whether he failed the second condition.

Now consider clumsiness. On any recognizable account of control, clumsy actions are uncontrolled, in some ways, to some degrees: that’s what their clumsiness consists in. *Yet some of them are intentional actions; successful ones even*. I think the basketball example Joshua gives us is an excellent example of this. When you look at *really* clumsy, yet successful, actions in slow motion, any success might look peculiar, unreliable, unrepeatable, unlikely, and so on. You might yell “Oh come on!” But that needn’t eliminate the idea that in some broad sense, what happened was *no accident*, given the agent’s powers and their exercise in these circumstances, and that what happened makes sense in light of what the agent wanted, perceived, believed and intended. For example: it was no accident that the basketballer threw the ball *towards* the net, rather than *away* from it. He did that because he *knew* where the net was, where he was, and so on. It *wasn’t* like entering a lottery. (Presumably a different sense of “lucky” is in play.)

So control’s exercise, in Joshua’s sense, is not a necessary condition of intentional action, because there can be surprising, rare, one-off cases of lucky, clumsy, uncoordinated, yet *successful*, and *intentional* action. Maybe they couldn’t be the norm. But that doesn’t mean they’re impossible.

Joshua might say the basketballer case is a case of rare success (40), where the agent puts themselves in a position to succeed intentionally, but the success itself must be attributed to luck, and not the agent. That seems implausible to me. Our clearest understanding of this contrast comes from considering the difference between entering and winning a lottery. Shooting *randomly* would be like entering a lottery, but this basketballer *didn’t* shoot in a random direction. In any case, if Joshua says this, then his reading of the end of his high school basketball career can’t be right. Cases of rare success don’t involve deviant causation, and they do involve control, restricted to putting oneself in a position.

I am not even sure that lotteries are a good guide in this context. Consider a would-be murderer, also suffering from late-stage Parkinson’s disease, shooting clumsily, spasmodically, and blindly into a dark room where they hope their victim is. This agent is like someone who keeps trying to buy lottery tickets, hoping to win, except the ticket machine doesn’t work reliably. But there are also crucial disanalogies. For one thing, it won’t be anything but *murder in the first degree* if this agent kills their victim. I’m pretty sure this agent does not *murder intentionally* on Joshua’s account, even if they kill their victim, because they don’t have flexible repeatability, their *operative* mechanisms are unreliable, etc. (At least: I’ve tried to construct the example that way. If it’s not quite there, add some conditions.) Yet one might well want a conception of intentional action that makes *this* intelligible as a *paradigmatic* case of *intentional murder*, in the first degree, should they kill their victim, regardless of whether the causation is a bit screwy and unreliable and not a good basis for manipulation or making predictions.

In raising these counterexamples, I am not really interested in counterexample tennis (although that’s fun sometimes). I’m really interested in whether control is prior in the order of understanding to intentional (or voluntary) action, as the structure of Joshua’s book suggests. Joshua quotes Jesús Aguilar (29) approvingly, where Aguilar says that cases of causal deviance all have in common undermining of agential control, and responses boil down to efforts to restore agential control. This suggests a biconditional connection between control and non-deviant causation. I am suggesting that one direction of the biconditional fails. Exercising a sufficient degree of control over behavior in the service of one’s ends might well be (part of) a sufficient condition for intentional action. But there might be rare cases where one causes something “in the right way”, *which doesn’t require that much*, without thereby exercising “sufficient control” in Joshua’s more-than-minimal sense. In that case, control is not prior in the order of understanding here: non-deviant causation is (if anything).

So here are some questions for Joshua. Am I right that, as applied to cases, he wants much more than the minimal interpretation of “sufficient degree of control”? Can he have that and have his reading of the basketball case? And more broadly, is the concept of control really prior to the concept of intentional action, or is the weaker, more minimal generic notion of “causing in the right way” prior?

*2. Explanatory Goals.*

Early on, Joshua says (3) he’s doing mildly revisionary conceptual engineering, aimed at “fruitful theorizing”, so we shouldn’t expect his definitions to line up exactly with intuitions or common usage, even if they are somewhat constrained by those. So one might object that the concept of intentional action I deployed in my argument just now is not *Joshua’s* conception. Fair enough. I think this raises important clarificatory questions about the title and spirit of the book, and its explanatory goals.

Joshua doesn’t say much explicitly about what “fruitful theorizing” is, but here’s my guess: for him, fruitful theorizing is scientific, and primarily good for *prediction, manipulation,* and *causal explanation*, given one’s interests in that regard, and secondarily, for *theorizing about the internal structure of agents*. For example: if one’s primary interest is to predict and manipulate how batters perform, and a batter performs notably badly when subject to jeering, it’s natural, secondarily, to posit an internal “public shame” factor that somehow causally explains the difference in success rate. If the batter can overcome the problem through practice of pre-game meditation, then it’s natural to posit some way meditation causally interacts with the shame factor. And so on. Throughout, fruitful theorizing concerns building accurate causal models of agents, given one’s interests, and accurate causal models of their internal structure, to the extent that’s relevant to one’s interests. So I suppose that’s another question for Joshua: am I right about “fruitful theorizing”?

If I’m right, then I wonder to what extent Joshua thinks alternative explanatory projects, governed by alternative conceptions of “fruitful theorizing”, are bankrupt, or misguided, or just different. For example: consider a broadly Anscombean account of the murderer I described before, where what they do manifests their practical knowledge of what they’re doing, and manifests, admittedly in an intermittent, unreliable, and mechanically very faulty way, their “knowing their way about” shooting and killing and things like that. On such a view, you could classify the case as a clear case of *intentional murder*, as I don’t think you can on Joshua’s view. (Note that it doesn’t really matter whether what happens is quite spasmodic and uncontrolled: that’s just how it is with Parkinson’s, and anyone who has Parkinson’s knows as much.) You might really want a definition of intentional action with this result, because you think that accommodating our legal and moral *knowledge* about the case, and the intended rational order it exhibits, is more important for “fruitful theorizing” than building predictively useful causal models of agential systems (even if you admit that, broadly speaking, agency is a causal phenomenon.) That is to say: scientists do not have the final say in *what it is* we study as philosophers of agency, and *what* they model when they build (very useful) causal models.

Another way to pose this kind of question about explanatory goals is to ask Joshua about *why* his book has the title it does. Joshua’s book is called *The* Shape *of Agency*. He didn’t choose the title *The* Nature *of Agency*. I think that’s very important to the spirit of the book. As Joshua says early on: “I am concerned with broad structures. I am less concerned with the skeleton [of agency], than with the shape of the skeleton.” (3) I *think* this means that Joshua is primarily concerned with the shape of certain *useful causal models*, rather than the (dried-up) skeleton of so-called “deep metaphysics”.

Let me illustrate by looking to Joshua’s account of control’s possession:

[Agent] J possesses control to degree DR with respect to some level of content approximation L for a plan-state P in circumstances C if and only if

[i] J’s success-rate at reaching L across C is DR

[ii] C is well selected, and

[iii] P plays a causal role in J’s behavior in every member of C.

Here possessing control implies, and is implied by, possessing a distribution of success-rates relative to a plan-state and a well-selected set of circumstances. A set is well-selected when “we follow principles for set selection that roughly mirror principles for building an accurate causal model of the agent as embedded in a broader causal system that comprises the kinds of circumstances in which we are interested.” (16) But this biconditional doesn’t tell you what’s *responsible* for this distribution of success-rates. Control is not appealed to *here* as something prior to the distribution which explains it; you can tell because the word “because” doesn’t figure in the biconditional. (I’m pretty sure this biconditional is not a *definition* of control’s possession either, as should become apparent soon.) Yet it is clear that Joshua thinks of *this* as his primary contribution in the chapter on control’s possession.

So what kind of “account” of control are we given here? Here control is a *measure* obtained by looking at an accurate causal model of how the agent behaves in some circumstances. It’s not itself a causal factor *in* a causal model, like a “shame factor” would be if it made a difference to a batter’s success-rates. Even if it were, the further question of *what it is* would be a *further* question.

Now later Joshua considers the objection that he hasn’t given a metaphysical account of what control *is*, so much as a behavioral profile which control must explain (21). His response is interesting. First, he says that metaphysically speaking, control is a set of causal properties sufficient for the relevant behavioral profile. (So we must distinguish control as *measure* of flexible repeatability from control *itself* which explains this feature.) Joshua does not further specify *which* causal properties are sufficient, which I initially found frustrating, until I thought more about what he’s really trying to do. Joshua doesn’t say more because he doesn’t *need* to in order to sketch the structure of a useful causal model. Second, he argues that if you really want a “deeper” metaphysical account of those properties, you’d do better thinking of them as dispositions-in-circumstances than as abilities. But he also explicitly disavows committing to any particular account of dispositions(-in-circumstances), again, because he doesn’t need to in order to sketch the structure of a useful causal model. Third, he acknowledges that this appeal to dispositions-in-circumstances is not really explanatory except in one very thin sense (28). But that’s not a big problem, because *that* kind of explanation is not his main concern.

This kind of abstraction crops up many times in the book. For example, there is no specific account of plan-states given, beyond saying intentions count (if they exist, which is for Joshua an empirical question), and noting (11-12) some extant literature that places constraints on what plan-states could be. I think that’s because, again, Joshua doesn’t need a specific (contentious) account of plan-states in order to sketch the structure of a useful causal model of (plan-directed) agency. It’s a feature of causal models popularized by folks like Judea Pearl, and interventionist accounts of causal explanation defended by folks like James Woodward, that to a large extent they *abstract away* from questions of so-called deep metaphysics. For most causal models, you don’t need to know *what* a “shame factor” *is*, metaphysically speaking, to give it a role in the model. It could be a state, relation, event, habit, process, fact, disposition, ability, mechanism, or complex of these, and still be well represented by a dot in a causal diagram. For the purposes of usefully modelling causal systems, including agents, one can to a large extent abstract away from finicky philosophical details about *what* one is modelling.

I could imagine someone complaining that this approach to the philosophy of agency dodges the “real philosophy”, because it abstracts *too much* from questions about *what* is being discussed. But I think that kind of objection is not very interesting, especially to the extent that Joshua’s book belongs to a broad naturalist tradition of thinking of philosophy as mostly continuous with empirical science.

I do think there is a related problem that is interesting and important. It concerns Joshua’s self-conception in the book as *solving* certain philosophical problems rather than *moving past them*.

One quick way to express this is to say that I suspect the book could be renamed *The Shape of* Causal Models *of Agency* without losing a great deal of content and interest. The models whose structure is sketched by Joshua could retain their use while abstracting from most details concerning *what* is being modelled. If one wanted to, the connection between these models and the nature of agency itself could then be filled in by commitment to some kind of scientific realism. But Joshua doesn’t argue for that in the book, nor should he, if his primary concern is sketching useful models that can abstract away from questions about relative merits of scientific realism, constructive empiricism, and similar views.

*3. Causal Deviance.*

I think the best way to illustrate the problem of self-conception is to have a quick look at Joshua’s claim to have *solved* the problem of causal deviance, rather than just moving past it.

Causal deviance is often understood as a problem for reductive accounts of intentional action, which say that intentional action is some X that is caused in the right way by some Y. The problem is to specify X, Y and especially “causing in the right way” without surreptitious appeal to the object of reduction. Davidson, for example, suggested “bodily movement” for X and “rationalizing belief and desire” for Y, and despaired of specifying “causing in the right way” without circularity.

Rowland Stout (2010) has argued that causal deviance really points to more general problems for philosophical views based on Humean accounts of causation, and that Aristotelian accounts do better. John Hyman (2015) argues for similar claims. The rough idea is that for *any* process-specific phenomenon, where there’s a distinctive *way* something is caused, it’s hard to specify “causing in the right way” in a counterexample-proof way if all you have to work with is two distinct existences (two events, say) and an arrow representing a causal relation between them. The reason is that the arrow in a Humean metaphysics of causation is indifferent to the difference between deviant and non-deviant causal chains, so it could always in principle be expanded to construct a deviant case as a counterexample. That’s a bit quick, but I don’t need to elaborate much to make my point. My point is that Stout and Hyman have two fairly well-known discussions of causal deviance which explain its philosophical significance in terms of the metaphysics of causation.

Joshua claims to have solved the problem of causal deviance. If he had, one would expect him to have paragraphs arguing that he has not snuck in some appeal to an object of reduction somewhere in his account, and paragraphs discussing how his solution relates to deep metaphysical issues about Humean and Aristotelian approaches to the metaphysics of causation, and so on. But I’m pretty sure there aren’t any paragraphs like that. So I think he has either misunderstood or misrepresented himself.

For example: a crucial component of Joshua’s account of non-deviant causation is the idea of a comprehensive set of circumstances. This is a kind of well-selected circumstance, where the actual case being modelled plays a “baptismal role” regarding the modal space surrounding it. I found Joshua’s discussion of comprehensive sets a little hand wavy and hard to follow, especially given the central role it plays in the book overall, but that minor complaint needn’t detain us now. The main point is that on his account, one follows principles of causal model construction in selecting a comprehensive set of circumstances, just as one does for well-selected circumstances. And Joshua does not even try to argue that those constructing such models do not make surreptitious appeal to a prior understanding of intentional action. Nor, I think, should he. Those who go in for causal models usually also admit that one cannot build them without presuming a great deal of irreducible causal background knowledge. In this context, that presumably includes knowledge about whether one is modelling a deviant or non-deviant case, or whether other cases are clearly deviant or non-deviant.

Must scientists building useful causal models of agential systems *solve* the problem of causal deviance, and our right to claim causal *knowledge*? Not at all. Consider occasionalism, where God is the only cause. Occasionalism could be true, and causal models still be useful (albeit only by the grace of God). Or imagine a world where every causal chain is stifled by a jealous deviant demi-god, who then replaces the stifled causality with its own causality. This is a world in which, so to speak, *every* causal chain is deviant. But the models still retain their usefulness (albeit, only by the grace of the deviant demi-god.) I think Joshua has mistaken *solving* the problem for realizing that it *poses no problem for what “fruitful theorizing” (by his lights) most centrally involves*, so he can just move past it.[[2]](#footnote-2)

By way of closing, let me briefly consider another way into this issue of self-conception. Joshua starts the chapter on causal deviance by talking it up as a problem for causal theories of agency, and ends the chapter sounding justifiably proud and relieved to have solved the problem. But in the middle, he refers approvingly to Alvin Goldman (37), who thinks that “deviant causation is not philosophically interesting or philosophically challenging.” Why would one be proud *or* worried about solving the problem of causal deviance, if one follows Goldman on these points? The answer is that one wouldn’t, if one is moving past the problem, rather than taking it seriously, and solving it.

That’s what I have. I hope I’ve provided good material for us to discuss further. Thanks.

**Sarah Paul’s comments.**

*The Shape of Agency* is a highly engaging book, managing to wed austere philosophical abstraction with an intimate, personal voice. It is chock-full of references to cultural icons like The Big Lebowski, Megan Thee Stallion, and Jay-Z, which renders it appropriate for me to begin with a Jay-Z lyric of my own: “Standing back from situations gives you the perfect view” (“Anything,” 2000). As I understand it, the aim of this book is to achieve an illuminating view of the broad structures of agency by standing back from many of the ongoing particular debates about the metaphysics and explanation of action. The focus here is on the form and not the matter, where according to Joshua, the form consists in possessing and exercising control over one’s behavior – and ideally, over one’s practical reasoning – in service of successfully realizing well-constructed plans.

There are copious things to praise in this book. In general, I find the nested-doll structure of the account, on which we first identify minimal conditions for intentional agency and build up to different kinds of agential excellence, quite congenial. However, I’ll primarily attempt today to convincingly inhabit my role as critic. I’ll start with a very big-picture question I have about the way that Joshua conceives of the philosophical project here. As we all know, one of the most vexing aspects of action theory is that so much depends on the very first step, in which we try to carve out the topic of interest. The general difference between passivity and activity is far too broad to be of much philosophical interest as such. A major part of the task is to justify focusing on certain kinds of activity or aspects thereof that are particularly puzzling.

Traditionally, there are various ways of doing this. For instance, much of the seminal work on agency emphasizes the centrality of activity that is subject to a distinctive kind of *explanation*. It is activity that is subject to a certain sense of the question ‘Why?’ There are longstanding debates concerning whether the fundamental form of explanation here is reasons-citing, psychological, teleological, or “naïve” in Michael Thompson’s sense, and whether or not these explanations are in part causal. But a general conviction has been that the ‘Why?’ question either defines our topic or does much of the work in pointing the way. A second kind of approach focuses on activity in which certain kinds of awareness or knowledge are present. And a third approach emphasizes activity that is attributable to an *agent*, rather than to some part or mechanism thereof. Conceived of this way, the question is how to make sense of attributability: how could action could spring from something worth calling the self, and thereby amount to autonomy or self-governance?

A very striking feature of Joshua’s book is that it says almost nothing about acting for reasons, or rationalizing explanations of action. The ‘Why?’ question is not discussed; the book is entirely about the question ‘How?’ Nor does he think that we should focus only on activity in which the agent knows what she’s doing, as we will see. The idea of attributability to an agent is present here, but it is simply taken for granted in the account he gives of intentional action (though he says a bit more about it later in the book). Rather, the primary focus is on activity or behavior that manifests the possession and exercise of control. It turns out on Joshua’s view that we can understand intentional action – usually a stand-in term for “the kind of action that is of special interest to philosophers” – largely in terms of controlled behavior. More specifically, he argues that intentional action is controlled behavior that is caused by and sufficiently approximates at least some relevant aspect of a good plan. The book has a lot to say about what it is for behavior to count as controlled to some particular degree. In contrast, Joshua deliberately prescinds from offering an account of what plan-states are, requiring only that they: a) represent (series of) events, states of affairs, or whatever, as to be done, or eventuated (that is, should must set out a goal) (b) play a causal role in the production of (or, at minimum, attempts to produce) the thing to be done (that is, M should move the agent towards the goal) (c) qualify as a state or states of the agent, as opposed to some sub-system of the agent.

But almost nothing is said up to this point in the book about the question of what makes a state or package of states attributable to the agent, in part (it is suggested) because the science is ongoing. This gives the impression that we can gain a good understanding of what intentional action is without any discussion of why controlled behavior in service of a plan-state amounts to the *agent’s* control of her behavior, or why activity caused by such states would be the kind of thing that is subject to a distinctive kind of rationalizing explanation. Joshua does go on in Chapter Six to discuss the kinds of organizational structures in creatures of various levels of sophistication, in virtue of which they are subject to assessment by behavioral standards. Here we get a sense of the kinds of considerations that bear on whether plan-states are attributable to an agent, and as I understand it, the minimal threshold seems to lie somewhere between a paramecium and a wasp. Further kinds of sophisticated representational and inferential capacities contribute to agential excellence, but are not essential to our topic.

None of this is necessarily intended as a criticism. I do think it would be helpful, however, if more could be said to situate Joshua’s approach within the field of action theory more generally, and to elaborate on why we should be particularly interested in such a capacious notion of intentional action. The emphasis on control has the advantage of highlighting the continuity between creatures like the paramecium, the wasp, the great apes, and human beings (not to mention robots). But many (though certainly not all) philosophers of action have thought we should focus our energies on theorizing about those aspects of agency that set us apart from the rest – the connection between plan-states and reasons for action; our ability to answer questions about not only how we are acting, but why; questions about why plan-states “speak for the agent,” such that their guidance constitutes the agent’s governance, and so forth. In the end, I found myself unsure where the proposals of this book fit. Should we take the account of control offered here as one piece of the philosophical puzzle, ultimately meant to be fit together with further philosophical work about the distinctive features of plan-states as manifested in sophisticated rational agents? Or is it rather a bold rethinking of what the priorities of action theory should be, with the implication that other ways of framing the central questions are misguided?

Without knowing exactly whether and how the theory aims to make contact with some of these other debates, it’s difficult for the reader to assess whether the account is “fruitful,” as Joshua claims his theory should be assessed (as opposed to whether it accords with common usage). To see the point, let’s look at his account of voluntary control, which is one of the fruits the approach is said to bear. Here, Joshua does claim to be developing a notion that can be put to use in illuminating traditional philosophical puzzles such as whether we have voluntary control over our beliefs, whether moral responsibility requires voluntary control, and whether addiction undermines it. His approach is to begin with the more general account of control he offers in the first part of the book, which roughly boils down to the capacity to flexibly and repeatedly match behavior to one’s plans, to some degree, across a well-selected set of circumstances. Joshua then adds a “non-compulsion condition,” as he puts it, as well as a condition that the well-selected set of circumstances must contain no impediment to the acquisition of the relevant plan states. In other words, you must have control, nothing should stop you from forming the plan to exercise it, and nothing should force you to exercise it.

Of course, a substantive non-compulsion clause is not so easy to come by. We might well think that the bulk of the philosophical progress to be made lies in fleshing out this part of the account – what is the difference between compulsion and voluntary action? The puzzle about addiction seems to lie precisely in the grey area between these two notions, rather than in any unclarity about whether the addict’s behavior flexibly and repeatedly matches the plans that cause it; their behavior is clearly controlled in this sense. But all questions about what compulsion is are set aside here. Further, the addition of the non-compulsion clause is made without any argument for why we should think of voluntary control as being at odds with the kind of compulsion exhibited in Joshua’s classic example of being held up for one’s wallet at gunpoint. This is too bad, since it’s a very puzzling (though in many ways appealing) claim that’s worth unpacking carefully. There is a sense in which giving up one’s wallet at gunpoint seems to be involuntary, or at least a “mixed” action, as Aristotle would put it – you don’t want to do it. But in another sense, you do want to do it more than you want to get shot, and you control your behavior accordingly. In my view, it’s fine to use the term ‘voluntary’ to signal the absence of compulsion and coercion, but we should be explicit that what we are doing in that case is employing it as an *ethical* notion: it is incompatible with compulsion and coercion (and perhaps ignorance) because those factors tend to be excuses. And to define the notion in this way is already to be well on the way to taking a stand on some of the aforementioned debates. If ‘voluntary’ is an ethical notion defined partly in terms of a lack of excuse, for instance, then it would be surprising if it were not closely related to moral responsibility.

On the other hand, those who are interested in control over our beliefs are not generally concerned with a notion that is incompatible with compulsion or coercion. Indeed, if we could deliberately form beliefs in response to a threat, that would be a resounding *vindication* of the kind of voluntary control over belief that most philosophers doubt that we have. That kind of control is more akin to what Joshua terms “direct” control: can we bring it about that we believe that P without self-manipulation, i.e. doing something else in order to cause ourselves to have that belief? Joshua remarks that the notion of direct control may not be as important as it is generally thought to be, since “why worry if the process takes a few steps as opposed to one, if the control is the same?” But the answer, I think, lies once again in the connection to reasons. No one doubts that we can believe in one step on the basis of evidence, i.e. epistemic reasons. What we (perhaps) cannot do is consciously form a belief directly on the basis of practical reasons like “I’ll shoot,” and this observation is meant to reveal something profound about the nature of belief.

My intention is not to question whether Joshua has made a valuable philosophical contribution in offering an account of control in its various forms, or to diminish the importance of trying to solve the problem of causal deviance. I’m merely trying to bring out the way in which our evaluation of the fruitfulness of the account as a whole might depend on getting more clarity about the questions these notions are meant to help illuminate – as opposed to those that he thinks are simply mistaken – and about how to understand the placeholders that are left. In the case of intentional action, there are places in the book where I worried that setting aside the topic of action explanation might lead to talking past other views rather than meeting them head-on. For example, Joshua discusses a challenge from a non-causal view of agency, representing his own approach as in some ways opposed to that way of thinking. But the central claims of most non-causal approaches, as I understand them, cannot be separated from questions about the explanation of action. Few who call themselves non-causalists would argue that there is such a thing as uncaused behavior, or even behavior that isn’t caused by some cognitive activity we might as well call plan-states. As I understand it, they simply think that that’s not what we’re trying to explain when we explain our intentional actions.

Again, this is simply a request to hear more about how Joshua conceives of the big picture. I think it’s fair to say that the book is written in a winkingly immodest tone, which can make it sound as though this is meant to be a more or less complete depiction – if abstract and architectural – of the outlines of agency, with the remaining details largely to be filled in by modern science. If that’s the case, then I disagree; I think a lot of important philosophical work is occupied by the placeholders here. But I doubt that what I just said is a good characterization of how Joshua conceives of the project, so I’d simply like some elaboration.

Now, lest I fall prey to the cardinal sin of commenting only about the foundations and ignoring the details, let me ask specific questions about two parts of the overall account. I’ll begin with the end, where Joshua offers a fascinating discussion of knowledgeable action. In contrast with some well-known views, Joshua denies that knowing what one is doing under some description is constitutive of acting intentionally. Rather, he depicts knowledgeable action as a form of agential excellence. However, he agrees with many of those same approaches that our knowledge of our own intentional actions is epistemically distinctive. Joshua likens it to other forms of self-knowledge and characterizes it as both peculiar and privileged, in Alex Byrne’s terminology: we come by it using a method that can only be used by oneself, and which results in beliefs that are more likely to amount to knowledge than those arising from alternative methods. To vindicate this idea, he offers a novel, practical-reasoning based account of the epistemic grounds of our knowledge. Roughly, the claim is that we are entitled to the judgments we make in the course of practical reasoning when and because we have sufficient control over that reasoning. One has sufficient control over one’s practical reasoning when one’s reasoning, in the relevant circumstances, reliably finds ways to succeed in action, or if the circumstances are unfavorable, delivers the best available alternatives. Thus, when an agent is A-ing and judges in the course of controlled practical reasoning that she is A-ing, her judgment amounts to knowledge. Joshua observes that these judgments may be *informed* by the agent’s intention, by perception, and by her experience of acting, but claims that the usual epistemic credentials provided by these sources “do not transfer,” as he says. Rather, they are credentialed by the agent’s control over reasoning and behavior.

There are aspects of this intriguing account that I’d like to understand better. First, it would be helpful to hear exactly what it is that Joshua thinks we know in this way about what we are doing. In discussion of others’ views, he mentions the idea that imperfective act descriptions exhibit a kind of “broadness” and “openness” – you can be in the process of A-ing, and know that you are, even if you are not now making any progress and will never actually succeed in A-ing. Some think that our practical knowledge is limited to these kinds of imperfective descriptions. But others, including many who Joshua takes to be his interlocutors, do not limit their claims in this way. Anscombe seems to think that the architect who has never seen or heard anything about the house has knowledge of what the completed house is like. And Imogen Dickie thinks that Robin Hood knows he is hitting the target with his arrow. Which side does Joshua fall on?

I think there is a dilemma lurking here. Suppose that the account is meant to vindicate the strong claim that when we are acting intentionally and our practical reasoning is sufficiently controlled, we know that we are successfully A-ing and will A. These judgments are claimed to be embedded in the agent’s ongoing practical reasoning. But Joshua characterizes practical reasoning as “directed towards figuring out whether one is succeeding, or whether one can succeed, or what the right answers to one’s questions about whether and how to act might be.” We engage in it “because uncertainty arises about whether what one is doing will lead to success, or whether one’s plan for how to do so will lead to success.” If this is right, then knowledge that we are succeeding or will succeed seems to be the *conclusion* of practical reasoning and not somehow embedded in it. If we already have such knowledge, the questions at which practical reasoning is aimed have already been answered. Further, to the extent that practical reasoning answers these questions, it’s hard to see how it would answer them on the basis of practical reasons. Joshua’s characterization of practical reasoning makes it sound a lot like theoretical reasoning about the topic of whether one is succeeding, and the kinds of information we have about that would seem to be of the ordinary empirical sort. On the other hand, if the knowledge we are seeking is only under a fairly non-committal, imperfective description, I have trouble seeing what controlled practical reasoning is adding. Why can’t we have that knowledge simply through some combination of intention, perception, and agential experience? Why would our *reasoning* need to be reliably successful in order to credential these judgments if the *truth* of those judgments doesn’t depend on actually succeeding in our actions?

We might also try framing the dilemma in terms of guidance. It seems to be central to Joshua’s account that the knowledge in question helps to guide reasoning and thereby plays a role in the rational control of action. But all he says about this (I think) is that judgments about what one is doing and how “set the background against which one reasons. One is reasoning about the best way to A, or about whether to continue A-ing. Such reasoning embeds judgments that one is A-ing, and often embeds judgments about how the A-ing is going at a time, at a relatively fine grain of detail.” But again, surely the judgments about how the A-ing is going at a time, at a relatively fine grain of detail, are ordinary empirically-formed judgments. It’s hard to see how they would be useful guides if they weren’t based on evidence. And if they are based on good evidence, why should we think that the usual epistemic credentials don’t transfer to this context? On the other hand, for answering the question about what the best way to A is, it doesn’t seem necessary to know that one is already successfully A-ing and will A eventually; isn’t it enough to know that A-ing is what one intends to do, or is trying to do? Indeed, knowledge produced by reasoning about the best way to A in some circumstance doesn’t seem to be privileged or peculiar, since I can reason about that just as well as you can even if you’re the one who is A-ing and not me.

I’ll end with a quick question about Joshua’s treatment of causal deviance: a pesky problem that Joshua claims to have solved for us. Essentially, for causation to be non-deviant on his view, the relevant behavior must be produced by way of causal pathways that reliably lead to behavior that approximates the content of the relevant plan, or parts thereof. We understand the notion of reliability, or “normal operation,” by reference to a comprehensive set of circumstances that are relevantly similar to the actual happening

I found it difficult to think through this abstract proposal, and I would have been helped by working through some of the canonical examples to see how it applies. In trying to do so myself, it seemed to me (perhaps mistakenly) that most cases of what are usually classed as antecedential waywardness, in which the deviance arises between plan-state and the ensuing bodily movement, will end up counting as non-deviant causation on Joshua’s view. Take Davidson’s climber. This case was discussed in the book, but many of the original details were altered or augmented in Joshua’s version (in Joshua’s retelling, for no apparent reason, the climber is depicted as much more dastardly than in the original – he actually intends to murder his friend and not just to save himself by letting go of the rope!), so I was uncertain whether what he says about the case was meant to apply to the original. But what he says is that if an agent’s plan is very simple, like “let go of the rope” – that is, without specifying precisely how she will do it – and if she is in circumstances where success is very easy to achieve, such that almost any movement she makes will cause her to drop the rope, then her dropping the rope turns out to be an instance of non-deviant causation. After all, the causal pathway from plan to behavior does reliably lead to success. Presumably, the same thing will be true of Frankfurt’s example in which a man spills his martini – for there is nothing easier than spilling one’s martini – and for Mele’s water-spiller. These examples all work because the relevant goal is so easy to bring about, a single tremor will suffice in most relevant circumstances.

There is a good chance I’ve misunderstood how the account works, and if so, I’d appreciate further clarification. But if not, hasn’t something gone wrong if we’ve reclassified many of the canonical cases of antecedential waywardness as at least potentially non-deviant? It seems to me as though no matter how simple the plan is and how easy success is to come by, the ensuing movement could either be caused deviantly or not. And this has to do with something like whether or not the movement is explained by the rationalizing properties of the content of the agent’s plan, as opposed to e.g. its unnerving properties. Joshua avers that that non-deviant causation is not a matter of the explanatory connections between content and behavior, but doesn’t say why he thinks this is the wrong place to look. In general, it would be very helpful to work through one or more of the standard cases I’ve mentioned and hear more about how and why the account treats these cases as it does.

I’ll stop there, and simply thank Joshua for a bold and interesting book that will give us all a lot to talk about.

**Joshua Shepherd’s responses.**

My thanks to Andrei for organizing, and to Kim and Sarah for their comments, which have helped me to see my own project more clearly. I feel a lot of gratitude towards Kim, Sarah, and Andrei: it’s a joy and a humbling thing to have one’s work read at such depth by such excellent philosophers.

A quick word about the nature of my responses: I do not have perfect words for everything, so some of the excellent points made will need to sit in the ether. In some cases I suspect Kim and Sarah have identified serious issues with aspects of my work, although I feel like more conversation is needed to bring out just what it is. What I will do, then, is respond directly to as many of the points and questions Kim and Sarah raise as I can. Roughly, I will talk about meta-level concerns first, the issue of control and intentional action second, non-deviance third, and knowledgeable action fourth.

Both Kim and Sarah raise concerns and questions about my methodology and about meta-level issues regarding my project. One issue that comes up for both of them is whether my theorizing is fruitful, as I claim, and what fruitful might mean for me. So, one issue Sarah raises is this:

Should we take the account of control offered here as one piece of the philosophical puzzle, ultimately meant to be fit together with further philosophical work about the distinctive features of plan-states as manifested in sophisticated rational agents? Or is it rather a bold rethinking of what the priorities of action theory should be, with the implication that other ways of framing the central questions are misguided?

Well, it is somewhere in between. I do not believe that I have rethought what the priorities of action theory should be. I do think, though, that some ways of framing the central questions are misguided, and I think that in this book I try to frame some questions that should be central – what is the nature of control? Of intentional action? Of skill? What’s the relation of knowledge to action and the role of knowledge in action? The results I offer can, I think, be integrated into further philosophical work on sophisticated rational agents. In fact that’s a topic I’m working on now.

A second issue Sarah raises is that in not focusing on action explanation what I say about action might be talking past the non-causalists. That might be fair. I am interested in the nature of intentional action, not directly in how we explain agentive behavior in all the explanatory contexts in which we do. Obviously there is some connection between these two issues, and there I suppose what I am doing is trying to remove some of the most pressing worries for the causalist side of things, rather than to take non-causalist views head-on.

Third, Sarah writes ‘I think it’s fair to say that the book is written in a winkingly immodest tone, which can make it sound as though this is meant to be a more or less complete depiction – if abstract and architectural – of the outlines of agency, with the remaining details largely to be filled in by modern science. If that’s the case, then I disagree; I think a lot of important philosophical work is occupied by the placeholders here.’

Right: I would disagree too. I am giving a picture of the outlines of agency. The hope is that this picture is satisfying in its own right, and answers some important philosophical questions. And a further hope is that this picture sets useful constraints on, and perhaps illuminates by suggesting avenues for reflection regarding, further theorizing about aspects of agency. I do not think the details are to be left to science, although I do think significant collaboration between philosophy and science will be required to get an understanding of the nature of sophisticated rational agents. I actually suspect philosophy should take the lead in that collaboration, but collaborations can be complicated.

Kim raises concerns in the same region, about how my theorizing might be construed as fruitful. In particular, Kim guesses that for me, ‘fruitful theorizing is scientific, and primarily good for *prediction, manipulation,* and *causal explanation*, given one’s interests in that regard, and secondarily, for *theorizing about the internal structure of agents*.’

This is partially right, although in writing this book I was probably not as clear about my own aims as I let on. But I think it’s not just prediction, manipulation, and causal explanation that I’m after – though those are useful things if we can get them. We need not be anti-realist to endorse the kind of methodology I take. For in offering explications (rather than analyses) of concepts, we might enhance understanding of target phenomena. This will be especially so if our pre-theoretic understanding of agency is deficient. Engineered concepts and webs of concepts might fly us closer to the sun.

Now, if one focuses on the first few chapters of the book, I think the complaint that I’m only offering causal models makes sense. Kim looks at my account of control and notes that all I really do is trace out the structure of control. And I think Kim is very perceptive about all this. So, when he says that ‘I could imagine someone complaining that this approach to the philosophy of agency dodges the “real philosophy”, because it abstracts *too much* from questions about *what* is being discussed,’ I think he is certainly right. I was also glad to hear him say that this kind of objection is not very interesting, given the fairly naturalistic tradition in which I am working.

Now, I don’t want to concede too much here. So, Kim says later that he reads my methodology as being ‘primarily concerned with the shape of certain *useful causal models*, rather than the (dried-up) skeleton of so-called “deep metaphysics”.’

I don’t want to be accused of thinking that deep metaphysics is dried up. It is more the case that my telescope seems to provide clarity at a certain level of theorizing. Where I see less clearly, I try to remain silent. Admittedly, I may have had too much fun implicitly or explicitly picking on certain ways of doing philosophy in the book. I may have been too winkingly immodest. That’s a problem I’ve had since grade school, and at times I find it a really annoying feature of myself.

But now look. Some features of my work here go well beyond causal models. There is, I think, a thread of normativity running through the book that is crucial, even if the normativity at issue is minimalistic.

Plans need to have some sort of structure – a means/end structure. That’s a normative constraint. It is true that control and non-deviance are relatively formal in the way I spell them out, and leave lots of room for a theorist to pick and choose what aspects of behavior are being examined for control or non-deviance. But once we level up to intentional action, I require that plans be good, where the goodness of a plan is its conduciveness to success given the goals the plan embeds. That’s a substantive claim, and one that has ramifications to which some will object. And then, in chapters 6-8, the normative notions begin to proliferate, and I think the result is that I am offering more than causal models, even if these models are a big part of the framework.

+++

Let’s turn to more specific kinds of worries. First, Kim argues that control is not necessary for intentional action. Along the way he discusses my notion of control, speculating that with my notion I want something substantive, something that rules out ‘*lucky, unusual, clumsy, uncoordinated actions* from being genuine intentional actions.’ (2) I do want something substantive, but I am not sure I want to rule these kinds of things out. As far as I can tell, one can perform actions in unusual, clumsy, and uncoordinated ways while exercising some degree of control. Luck is probably a different matter. While I wouldn’t rule all operation of luck out of intentional action, in some sense luck is opposed to control, so I would rule out overly lucky behavior from qualifying as intentional action.

Now, consider the clumsy basketball player. Kim claims the following:

‘When you look at *really* clumsy, yet successful, actions in slow motion, any success might look peculiar, unreliable, unrepeatable, unlikely, and so on. You might yell “Oh come on!” But that needn’t eliminate the idea that in some broad sense, what happened was *no accident*, given the agent’s powers and their exercise in these circumstances, and that what happened makes sense in light of what the agent wanted, perceived, believed and intended. For example: it was no accident that the basketballer threw the ball *towards* the net, rather than *away* from it. He did that because he *knew* where the net was, where he was, and so on. It *wasn’t* like entering a lottery.’ (3)

I agree with all this. It might be helpful to keep control, as a measure of reliability relative to some standard (a means-end aspect of a plan) and some model of the circumstances, distinct from intentional action. Intentional action requires control but allows significant latitude in what amounts of control are required in any given case. My own judgment about this case is that making the shot would not be an intentional action, since the agent had so little control over behavior under this description. But aspects of what the agent did could be described as intentional action.

But Kim is stalking a broader point. He is ‘really interested in whether control is prior in the order of understanding to intentional (or voluntary) action,’ and he claims that ‘Exercising a sufficient degree of control over behavior in the service of one’s ends might well be (part of) a sufficient condition for intentional action. But there might be rare cases where one causes something “in the right way”, *which doesn’t require that much*, without thereby exercising “sufficient control” in Joshua’s more-than-minimal sense. In that case, control is not prior in the order of understanding here: non-deviant causation is (if anything).’

From the official stance of what I wrote in the book, I don’t separate the exercise of control from non-deviant causation, so I don’t need to disagree that non-deviant causation is prior in the order of understanding to intentional action. But to say that is just to say that the exercise of control is as well, and therefore also the possession of control, since that notion is key to understanding exercise.

But Kim seems to be driving at a different idea when he talks about my more-than-minimal sense of control, and about a notion of non-deviant causation that is minimal and generic. So, what might count as a case of non-deviant causation in the absence of control? I’m open to thinking further about this. Kim offers one kind of case:

Consider a would-be murderer, also suffering from late-stage Parkinson’s disease, shooting clumsily, spasmodically, and blindly into a dark room where they hope their victim is. This agent is like someone who keeps trying to buy lottery tickets, hoping to win, except the ticket machine doesn’t work reliably. But there are also crucial disanalogies. For one thing, it won’t be anything but *murder in the first degree* if this agent kills their victim. I’m pretty sure this agent does not *murder intentionally* on Joshua’s account, even if they kill their victim, because they don’t have flexible repeatability, their *operative* mechanisms are unreliable, etc. (At least: I’ve tried to construct the example that way. If it’s not quite there, add some conditions.) Yet one might well want a conception of intentional action that makes *this* intelligible as a *paradigmatic* case of *intentional murder*, in the first degree, should they kill their victim, regardless of whether the causation is a bit screwy and unreliable and not a good basis for manipulation or making predictions.

This case connects with some big-picture questions Kim has about how I see other explanatory projects in the philosophy of action. There’s a lot here that I can’t touch. So let me just say: I do not think that moral judgments should guide theorizing about the nature of intentional action. That over-moralizes the notion, it seems to me. One legitimate theoretical aim is to classify the kinds of behaviors that call for moral sanction, what holds them together, and so on. I’m not sure why the term intentional is the go-to in this case, since we have other options. Maybe morally responsible behavior departs from intentional action in some cases, without much loss to the degree of responsibility we attribute.

But I’m not inclined to fight too hard over possession of the term ‘intentional,’ and I am open to a theoretical project that places practical knowledge as the center, rather than control. I agree with Kim when he says that ‘scientists do not have the final say in *what it is* we study as philosophers of agency, and *what* they model when they build (very useful) causal models.’ And I do not think a project that centered practical knowledge would be directly at odds with my own, though there would be disagreements. Comparing the convergences and divergences between, for example, what such a project would say about skill and knowledge, or about knowledgeable action, or about moral skill, might itself be a fruitful exercise. (Cross-cutting schemes of classification is not necessarily bad in science. And it is not obvious to me that cross-cutting in philosophy is necessarily detracts from understanding.)

+++

Now let me turn to worries about causal deviance, raised in different ways by Sarah and Kim.

There’s an interesting feature of deviance. For some, it is completely ignorable. For others, it is minor. Sarah describes the problem as ‘pesky.’ Gnats are pesky; bears are not. For others, it is major. In many papers it is at least cited as a major flaw in causalist accounts and a major motivation for pursing some alternative program. So, I felt there was adequate motivation to deal with it for this reason – to render such appeals irresponsible short of taking seriously the ways it can be dealt with.

Sarah points out that in many cases of deviant causation – Davidson’s climber, Mele’s water spiller, Frankfurt’s martini spiller – it looks like success is very easy to achieve. And if there is a reliable, because easily found, pathway from plan-state to success, don’t I classify these as cases of non-deviant causation? And, quoting Sarah, ‘hasn’t something gone wrong if we’ve reclassified many of the canonical cases of antecedential waywardness as at least potentially non-deviant?’ Sarah goes on to say the following. ‘It seems to me as though no matter how simple the plan is and how easy success is to come by, the ensuing movement could either be caused deviantly or not. And this has to do with something like whether or not the movement is explained by the rationalizing properties of the content of the agent’s plan, as opposed to e.g. its unnerving properties.’

In response, Sarah is right that I classify cases of easy success as cases of non-deviant causation. But there is a potential slide here that I want to highlight. My cases of easy success involve full specification of relevant causal parameters. Traditional cases of antecedential waywardness leave deviance to intuition. The way that these cases are articulated, they do not look like cases of easy success to me. When we imagine the martini being very easy to spill, we compare the spill case with other cases that involve a more controlled manner of spilling – one that, perhaps, was in the plan’s content. So the jittery spill seems deviant. In a legitimate case of easy success, there is no particular manner of spilling specified in the plan, however, so the way the success is brought about is not legitimately described as more controlled or less controlled. Since martini spilling in the real world is often a bit more complicated than a constructed case of easy success, imagination tempts us to imagine circumstances that are out of bounds.

How does this square with Sarah’s distinction between the rationalizing properties of plan content, and the unnerving properties of plan content. I’m tempted to functionalize this distinction. So I would insist that in cases of easy success the success is explained by the rationalizing properties. To get a grip on the difference, there must be a difference in behavior between the rationalizing properties and the unnerving properties. If unnerving properties and rationalizing properties behave the exact same way no matter slight differences in circumstances, then the distinction fails to get a grip.

+++

Kim’s worries about non-deviant causation are different. He suggests that I do not solve the problem, but merely move past it. I admit that’s not a distinction I’m familiar with, and so I’ve had a long think about this, and about what can be legitimately demanded of someone who claims to solve a philosophical problem. One thought I’ve had is that, putting solving to one side, there are better and worse ways of moving past a problem. The incredulous stare might be one way to try to circumvent a problem. It is usually not a great one. But sometimes someone demonstrates how to think about a problem in a way that renders it unproblematic. That might be a virtuous way of moving past a problem.

What would count as solving the problem? Kim suggests Stout and Hyman’s discussions of deviance, ‘which explain its philosophical significance in terms of the metaphysics of causation,’ are examples of what would count.

Now, my reading of, for example, Stout, is that we cannot solve deviant causation unless we adopt a non-Humean metaphysics of causation, one that involves identifying mechanisms that work by way of processes that are the realization of the mechanism’s potentiality. There is a lot of technical language here, and work to do to spell out how events or results belong to one process rather than another, how to think about the realization of potentiality, and so on. So this is more like the beginning of a solution. But Stout claims things like on a process model of causation, ‘There is no space in the process of a mechanism being realised in which a deviant causal chain can be interpolated. If a deviant causal chain were in place then the result would belong to the operation of a different mechanism.’ I’m dubious, given that I don’t think we could understand every action as the product of a single mechanism. Often we have a concatenation of mechanisms, that need to be coordinated, and therein lies a possibility for deviance. But my bigger worry, here, is similar to the one I put to Wayne Wu in the chapter: there is an attempt here to find a kind of magic, in this case a special feature of causation in virtue of which causation would be non-deviant. That’s looking in the wrong place. One way to put the response, then, is this: I offer no solution to deviant causation because there is no solution, thus conceived. The right way to solve the problem is to understand how to move past it.

+++

I turn, last, to Sarah on knowledgeable action. Sarah asks for clarity regarding what it is we know about what we are doing in the way I specify. Is it just the rough-grained imperfective description, or is it something more detailed? I would say that it is the latter. This opens me to the second side of a dilemma Sarah lays out. The worry, if I understand it correctly, is that this more specific knowledge of what we are doing seems to have knowledge that we are succeeding as the conclusion of practical reasoning, not as embedded in practical reasoning as I claim. A further problem, I think, is Sarah’s worry that practical reasoning, under my treatment, looks suspiciously like theoretical reasoning about whether we are succeeding, and if that is right the knowledge we acquire does not seem so special – it seems based on standard non-practical sorts of evidence.

I suppose I would say, in response, that we don’t know that we will succeed, not always. What we know is something like where we are at in a spatiotemporal map that the plan for action lays out, how that relates to where we are supposed to be, and whether we are thereby succeeding. In generating knowledge of where we are at, and how and why we are proceeding as we are, the practical reasoning guides the execution of the action, in part by directing the next moves within reasoning, as the agent keeps the plan updated to the needs of the moment. Knowing what we are doing at a time, and to what degree we are succeeding, can lead to other patches of reasoning about how to do what is next. So, while much of the knowledge we have about how things are going relies upon standard sorts of evidence, the agent’s control over practical reasoning is at the leading edge of this knowledge.

I’m not sure if those remarks are sufficiently responsive to Sarah’s worries, so let’s have a go at the second way she frames the dilemma, in terms of guidance. On one hand, do we really need more detailed knowledge regarding where we are at in the progress of the action? Sarah says that all we might really need is knowledge of what we are intending or trying to do. I suppose I don’t feel the pull of this. For at least some actions knowledge of intention or of trying seems to fall far short of informing the next best move. So perhaps I’m on the other side of the dilemma again. And here Sarah wonders whether there is anything practical about the knowledge I have in mind. For, quoting Sarah, ‘surely the judgments about how the A-ing is going at a time, at a relatively fine grain of detail, are ordinary empirically-formed judgments. It’s hard to see how they would be useful guides if they weren’t based on evidence. And if they are based on good evidence, why should we think that the usual epistemic credentials don’t transfer to this context?’

In response, let me say that for a long time, in part because of Sarah’s work, I was convinced that non-observational knowledge was something of a red herring. And I still wouldn’t deny that some of the knowledge we have of action is observational or ordinary in nature. In fact I now think there is an interesting interplay between observational and non-observation knowledge in some actions – for example in improvisational actions. So let me ask: is there room for any non-observational knowledge in knowing how the A-ing is going at a time? My thought, in the book, is that there is because the control over practical reasoning provides a different sort of entitlement for the judgments made while reasoning practically. The entitlement derives from the control that guides the action, not from whether the evidence warrants this or that judgment. So I allow for there to be some space between what the agent’s evidence indicates regarding what the action is and how the action is going, and the agent’s judgments regarding all this. The agent has a latitude provided by the space of behaviors she can control, such that she can work her action out in ways the third-personally accessible evidence may not indicate.

+++

1. Initially (64) we are told only that intentional actions imply the exercise of control; over the page (65) we are told it is a sufficient degree of control that is required. See (113) for the account of degrees of control. [↑](#footnote-ref-1)
2. [If time, mention Megan Thee Stallion “Money good” quote here. Causal models have a role in making predictions, bets, and money. (Moneyball). Lyrics about “embracing the fake” seem apt here: that is a *challenge* to pretentious critics.] [↑](#footnote-ref-2)