# 6

# Pluralistic Attitude-Explanation and the Mechanisms of Intentional Action

*Daniel C. Burnston*

## 6.1 Introduction

According to the Causal Theory of Action (CTA) (Bratman 2000; Davidson 1963), a bodily movement is an action if it is preceded and caused by a reason that the agent had for so-acting. Reasons, on the same theory, are reducible to complexes of propositional attitudes, in particular the states posited by common-sense psychology, such as beliefs, desires, and intentions. An 'agent' is a being whose actions are caused in the right way by their attitudes (Schlosser 2011).

The CTA, on my view, gets something importantly right. It is the mental states of agents that determine which of their movements are actions (Smith 2012). However, naturalizing the CTA is problematic. In this chapter, I will analyze the relationship between the CTA and current models of decision-making from the sciences. I will argue that there is no straightforward way to map attitudes onto those mechanisms.

First, I will distinguish between two perspectives on, or ways of pursuing, the CTA. The first is *implementation-neutral*, and treats the attitudes as explanatory kinds whose nature is to be uncovered by the relevant empirical investigations. On this view, attitude-explanation is flexible with regard to interacting with empirical data. The second is *implementation-specific*. This means that the CTA is to be tied to a particular construal of the nature of the attitudes, and their causal role in producing action. This perspective leaves the CTA, or at least particular versions of it, falsifiable by empirical data and theory.

The standard approach to the CTA in philosophy of psychology is implementation-specific. These views posit a separate 'practical reasoning system' that operates on conscious, abstract, conceptual representations

*Daniel C. Burnston, Pluralistic Attitude-Explanation and the Mechanisms of Intentional Action* In: *Oxford Studies in Agency and Responsibility.* Volume 7. Edited by: David Shoemaker, Oxford University Press.
 DOI: 10.1093/oso/9780192844644.003.0007

corresponding to the attitudes, and outputs intentions (Pacherie 2000; Carruthers 2004). Intentions then interact with the motor system, the role of which is to enact the content of the intention. I will suggest that there are both conceptual and empirical problems for this view, propose an alternative, and analyze what that alternative suggests about the explanatory role of the attitudes.

In Section 6.2, I will explain the two perspectives on attitude-explanation. In Section 6.3, I will explain the standard operationalization of attitudes like 'intention' in philosophy of psychology, and the problems that this view faces. In Section 6.4, I will outline an alternative view of decision and action, which posits that decisions are the outcomes of a competition process between 'event files' represented in perceptual and motor systems. In Section 6.5, I will outline a pluralistic perspective on attitude-individuation in light of these results, and in Section 6.6, I will assess the kind of control that deliberative agents have in light of the mechanisms discussed. Section 6.7 concludes.

## 6.2 Approaches to the attitudes

There are, broadly, two distinct perspectives we could take towards attitudes like 'intentions' or 'beliefs'. First, we could take our everyday practices of citing these mental states as *implementation-neutral*. That is, we could treat them as a kind we pick out in the world and converse about, but saddle those practices with minimal commitments about the structural properties and causal processes by which this kind operates. The second perspective, conversely, is *implementation-specific*. That is, we should interpret our practices as undertaking ontological commitments of a particular sort, and then hold our concepts accountable to how the world turns out.

Most philosophy of action does not say explicitly which perspective it works under when talking about the attitudes. One frequently comes across claims about the psychological functioning of the agent (see, e.g., Bratman 2001), suggesting that, at least broadly, the discussion of mental states incurs psychological commitments. But these claims are generally cast at a very abstract functional level. And just as often, the debate focuses on the structure of action explanation, and whether those explanations must cite causal explanations as opposed to some other kind of explanation—e.g., logical or teleological. It seems one can ask these questions without any detailed psychological hypothesizing. I will not attempt any kind of systematic

assessment of these perspectives in the literature, since my project here is to cash out the different perspectives and their commitments.

We can, however, see clear cases of each perspective at work. Consider Railton's (2012) approach to the attitudes, which is informed by psychological and neural evidence about reinforcement learning. Attitudes, according to Railton, help constitute *internal models* of the world, which are learned by the brain. Beliefs involve expectations about how the world will evolve, and desires about the motivational consequences of those outcomes. On this view, the *nature* of beliefs and desires are discovered in the functioning of the nervous system. Reinforcement learning processes operate subconsciously in widespread neural networks, and are shared between humans and other organisms (Cushman et al. 2017). Or, consider Wu's (2011) functional characterization of intentions. On Wu's view, intentions solve 'many-many' problems—i.e., mapping stimuli onto possible outcomes. This is a functional construal that leaves open the structure of the state and the causal details by which the function is implemented.

Conversely, consider Churchland's (1981) famous case for eliminativism. Churchland takes folk psychology to be a genuine theory, which is deeply committed to not only a characterization of the rough functional roles of the attitudes, but also to a view of their structure (sentential) and the kinds of processes in which they can figure (syllogistic). Churchland's motivation for eliminativism is his view that the actual (neural and psychological) processes that produce behavior do not instantiate those properties.

The two perspectives are thus quite distinct in how they view the relationship between attitude-explanation and empirical evidence. The implementation-neutral perspective is extremely *flexible* with regard to how to fit the attitudes into the empirical world. It presumes, only, that there is *some* natural kind that our attitude-explanations pick out. It does not presume that the kind's nature is revealed by our practices, that the kind exists only in the places we usually expect to find it (bees, for instance, may have beliefs and desires on this view), etc. As such, there is no particular causal structure that is entailed by our practices of attitude-explanation.

The implementation-specific perspective, on the other hand, demands that the attitudes have an operationalization that is testable in light of what science tells us about the mind and brain. According to Churchland, empirical facts about neuroscience simply *disprove* the everyday notion of propositionally structured attitudes standing in syllogistic relationships to each other. As such, those attitudes should play no role in our scientific or philosophical explanations.

Recent philosophy of psychology had taken an implementation-specific approach to the attitudes. States like 'intention' are taken to be realized in our heads, to have specific properties, and to play distinctive causal roles in the generation of action. According to these views, substantiating the attitudes requires operationalizing them in such a way that they play a role in a falsifiable theory of cognitive architecture.

I will discuss these views in detail below, and criticize them in the following sections. I'd like to be clear about my overall strategy, however, before we begin. I do not think we are compelled by any empirical or theoretical reasons to take the implementation-neutral or implementation-specific perspective to the attitudes (Sehon 2013). What I want to do in this chapter is suggest that more attention needs to be paid to the perspective being employed in particular instances, and to point out pitfalls with each option. If we go with the implementation-specific approach, then there is strong reason to think that the standard operationalization of the attitudes in philosophy of psychology is false. If we go with the implementation-neutral perspective, however, no single kind underlies our folk practices. Instead, they range over a variety of distinct mental states which play distinct roles in generating actions. At the end of the chapter, I will suggest that this should motivate us to adopt a kind of pluralism about attitude explanations.

## 6.3 Operationalizing the attitudes

The CTA construes tokenings of propositional attitudes as (i) preceding action, and (ii) causally determining action. In philosophy of psychology, this view is generally interpreted by construing decisions as the result of a *practical reasoning* system, which is distinct from, but interacts with, the motor systems that control bodily movements. It is coordination between the practical reasoning system and the motor system that implements the causal determination of action by propositional attitudes. This kind of view has seen several refinements in the work of Searle (1983) and Pacherie (2000, 2008). One of the refinements has been to note that there is a kind of intentionality in the motor system itself, which is involved in the control of movements and their orientation towards a goal. Seminal experiments by Fourneret and Jeannerod (1998) showed that subjects can monitor and adjust their goal-directed motor movements even when they are unaware that they are doing so.

In light of these results, most admit that the motor system can represent goals and control movements (Butterfill and Sinigaglia 2014). The distinction between the practical reasoning system and motor system is then construed in terms of the *abstract* and *distal* nature of representations in the practical reasoning system. On this kind of view, the practical reasoning system represents distal aims at an abstract level (e.g., 'drive to the airport'), while the motor system represents immediate goals and the movements required to reach them. 'Decisions' are the outcome of the practical reasoning system and produce intentions. Intentions then interact with the motor system via a 'content-preserving causal process' that determines which motor action is performed on the basis of the content of the intention, and thus rationalizes the movements (Mylopoulos and Pacherie 2017, 2018).

Propositional attitudes, on this view, play particular roles in action in virtue of their structure. They are taken to be good for deliberation and practical reasoning because of their abstract contents and discursive organization (cf., Fodor 1983). Motor representations, on the other hand, are good at motor control because they implement a different kind of content, namely representation of motor kinematics (e.g., flexion and tension in the muscles, limb position), and represent that content non-conceptually. This has become known as the 'format distinction' in the literature.

I will refer to this as the 'standard' interpretation of how the CTA relates to cognitive architecture, and it does have some support in cognitive neuroscience. It is common to think of brain systems as organized according to 'abstraction hierarchies' (Uithol et al. 2012; Uithol et al. 2014) with more abstract representations at the top of the hierarchy and more specific representations at the bottom. In the field of motor control, it is often assumed that a causal process runs from more abstract levels of the hierarchy to more specific ones. Here is an example Grafton and de C. Hamilton construe:

> Motor control as a refinement of information processing from a distant goal ('light the cigarette') to a more detailed motor plan ('lift the match, strike the match') to a precise specification of the reaching and grasping actions required to achieve each motor plan, and finally the activation of specific muscles in a coordinated sequence and the associated coarticulation that emerges at this level of organization.
>
> (Grafton and de C. Hamilton 2007, pp. 595–6)

In the remainder of the chapter, I will focus primarily on intentions. Intentions are particularly important on the standard view, because they are

the bridge between the practical reasoning system and the motor system. As such, they are *distinct* from motor representations, they are the *outcome* of decision, and they causally determine the action that the motor system undertakes.

There are two major problems for the standard view, one empirical and one philosophical. The empirical problem we can refer to as the 'localization' problem. Given that there is a distinct role posited by intentions on the standard view, as operating at the juncture of two distinct systems (the practical reasoning system and the motor system), one might expect that there are particular places and times in which the two systems interact. But more 'abstract' states that are causally prior to action have been located in *many* different neural areas, and at many different times prior to an action (Uithol et al. 2014). As such, there is no precise causal juncture between practical reasoning systems and motor systems. Speaking somewhat more broadly, for *any* abstract, decision-related variable that neuroscientists have tried to measure—including 'rules', 'plans', 'task sets', 'goals', 'values', and others, the neurophysiological corollaries of those variables are widely distributed across the brain, and deeply bound up with perceptual and motor information (see, e.g., Fusi et al. 2016; Kennerley et al. 2011).

While one might hope that continued investigation will specify the causal nexus between different putative systems, the standard view also faces a philosophical problem, the 'interface problem' (Burnston 2017a; Butterfill and Sinigaglia 2014; Ferretti and Caiani 2018; Mylopoulos and Pacherie 2017). On the standard view, intentions and motor representations have distinct kinds of contents—conceptual and discursive, versus non-conceptual and motoric. The view, however, still supposes that intentions *cause* and *control* specific motor representations. The interface problem is how states like intentions can exert this specific influence on the motor system, given the differences between their respective contents.

Consider the propositional intention to 'grasp the ice cream'. In virtue of its abstract contents, it is equivocal between all of the motor details involved in using a particular grasp to obtain a particular cone (or bowl!) of ice cream. But the actual grasping requires specification of those details. How, then, is a 'content-preserving causal process' that begins with one type of content supposed to eventuate in the other type? A variety of solutions to the interface problem have been proposed, and I will not discuss them in detail here. In my view, none of the solutions overcomes what I've dubbed the 'diversity and specificity' of the intention-to-motor representation relationship (Burnston 2017a). Even if you fine-grain the conceptual states

involved (e.g., 'grasp the ice cream *cone*' as opposed to 'grasp the ice cream'), there will be a diversity of possible motor representations corresponding to the intention. However, in each actual case of action, a particular motor representation must be tokened. The content of the intention has no resources to bridge the gap.

More deeply, I have argued that the interface problem is not a problem we *should* solve. If it weren't for the functional role posited for propositional intentions by the standard view—i.e., a state that is produced by a decision and in turn produces specific effects in the motor system—then the structural distinction between intentions and motor representations would not need to be bridged. As I will show below, there is a different characterization of decision and action on which the structural and functional characterizations come apart. On this view, the interface problem is not a problem at all.

## 6.4 An embodied view of decision-making

The alternative view of decision and action control that I will discuss involves two key components, *event files* and *thresholded competition*. It questions the key presumptions of the standard view, namely that (i) *decisions* are made in an abstract or amodal reasoning space, and (ii) the role of motor systems is primarily to carry out the previously made decision. Instead, the alternative view proposes that potential actions and outcomes are represented *in perceptual and motor systems*. It further proposes that, rather than being driven by propositional reasoning, decision is the outcome of a *competition process* between action possibilities represented in parallel. Sections 6.4.1 and 6.4.2 will discuss these aspects, respectively. In Section 6.4.3, I will argue that discursively structured representations do interact with this system, but that their role is to *bias*, not *determine*, the outcomes of the competition process. In Section 6.4.4, I will endeavor to show how this kind of approach can scale up to planning, deliberation, and economic reasoning. This will set the stage for the pluralist approach defended in Section 6.5.

In what follows I will broaden the discussion of the 'motor system' to include interactions between perception and motor representation. The reasons for doing so will become clear—namely, there are direct interactions between perception and motor representation that do not need to be mediated by concepts or propositional reasoning. As such, the current project

belongs under the aegis of 'grounded' or 'embodied' cognition, which I construe as the thesis that cognitively sophisticated functions are partially constituted by perceptual, motor, and affective representation. (This characterization is dissociable, and should be distinguished from, the extended mind and anti-representational views with which it is sometimes associated.) I will hence refer to the alternative picture as an 'embodied' view of decision.[1]

### 6.4.1 Event files

James (2013) held that the perception or imagination of an action's *outcome* automatically activates the motor representation of the action. This 'ideomotor' theory of action has been revived by Hommel (2013), among others, and given a representational basis. The idea is that learning produces bidirectional associations between perceptual and motor representations, and that these can then be triggered in either direction. The importance of this idea for the discussion is this: the standard story assumes that perception and motor systems, on their own, have no power to initiate action. Decision-making requires a separate system to select the appropriate action and initiate it. The ideomotor theory fundamentally denies this view—actions *can* be triggered by perception, without the need for a separate decision maker.

In the abstract, the methodology for investigating event files pioneered by Hommel is as follows. First, one has subjects perform a motor act that is arbitrarily paired with a particular perceptual outcome. So, for instance, pulling a lever may be associated with a light flashing or noise occurring. After training on the task, subjects then are given the perceptual outcomes as *initiating* stimuli for a required action. The relevant contrast is between conditions on which the required action is *compatible* with the previously learned association, or incompatible. Across a wide range of stimuli and action types, it has been shown that compatible conditions expedite performance compared with incompatible conditions. The conclusion is that a bidirectional association has been learned in the first part of the experiment, and is activated automatically by perception of the former action outcome. This in turn facilitates the associated action, relative to the one required in the incompatible condition (Figure 6.1).

[1] For related views on action, see Schurger and Uithol 2015; Sims and Missal 2019.

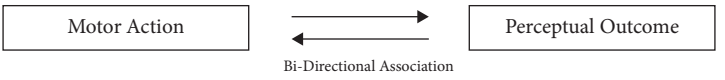


**Figure 6.1** Event files

Event files are rich representational structures. They encode information that is not relevant for the task in which they're learned (Hommel 1998). They can be extended in time and space, with outcomes spatially and temporally separated from actions (Stoet and Hommel 1999). And they are not limited to encoding simple perceptual stimuli—event files can include both linguistic stimuli and categorical visual stimuli as the perceptual aspect of the association (Waszak et al. 2003). They can also encode affective outcomes (Lavender and Hommel 2007).

The event files framework argues that there is an important relationship between an action and its *perceived outcomes*. Rather than having to reason propositionally about what will result should we undertake some action, these expectations are encoded via direct associations between perception and the motor system. But there is a front-end to this process as well. Objects in our environment contribute to the behavioral options we have available, and Cisek and Pastor-Bernier (2014) have argued that perception of these *affordances* is fundamental to the generation of action. Combined with the notion of event files, what we get from affordance perception is this (Figure 6.2): objects in our environment afford certain actions that lead to certain outcomes. The idea then, is that when we perceive objects, these perceptions trigger the potential actions associated with those objects, and that these in turn trigger associations with the expected perceptual outcomes of those actions.


**Figure 6.2** Event files and action affordances

### 6.4.2 Thresholded competition

In general, in a behavioral scenario, we are presented with a large number of objects and action affordances, which will lead to a variety of outcomes. How does the choice between outcomes occur? While the standard story

says that this selection must occur top-down from a distinct decision maker, there is an alternative, *competitive* model of decision-making. On the kind of view put forward by Cisek and Kalaska (2010), objects in our environment, and their associated action-affordances, compete for motivational salience. So, for example, the drinking-affordance of the water-jug is more salient when one is thirsty, and the eating-affordance of the bag of almonds is more salient when one is hungry. On the standard picture, one must first translate the perception of both almonds and water into beliefs, check these against one's desires (to drink or eat, respectively) and then form an intention as to which to consume. The competition account denies the necessity of this process—it argues that the decision is the outcome of a competition between the behavioral options, as represented in perceptual and motor systems, rather than encoded top-down by a distinct process.

Providing evidence for a competitive sensorimotor picture involves showing that multiple distinct possible actions are represented in parallel in sensory and motor systems. There is such evidence in both neuroscientific and psychological data. When a monkey is presented with multiple behavioral options, neurons in the premotor cortex begin to signal *both* possible actions *prior* to the cue signaling a particular stimulus as the correct action target. Moreover, these signals are modulated by the *subjective value* of the reward associated with each stimulus (Cisek & Kalaska, 2005).

Hence, there is evidence that the brain's motor systems are involved in value computation, but the way this computation affects decision is by influencing competition between actions represented in parallel in the motor system. Wolpert and colleagues (Gallivan et al. 2017) have performed a variety of experiments using 'go before you know tasks', in which human subjects must launch a reach movement towards a set of targets prior to a signal which clarifies which is the appropriate one. A robust finding is that the initial launch trajectory is towards the spatial average of the potential reach targets. Computing this average requires representing both reach trajectories in parallel.

In terms of perceptual information, Mante et al. (2013) have shown that non-chosen perceptual action targets are represented in neural population activity along with chosen targets. Models of decision-making in neuroscience strongly reflect this notion of competitive processing. In Loh and Deco's (2005) model of decision in the prefrontal cortex, for instance, multiple stimulus-action associations are represented in parallel, and the populations representing these associations compete with each other directly via mutual inhibition to drive behavior. Input from the environment serves to

bias this competition, and the population evolves until a threshold is crossed, after which the winning association drives behavior.

### 6.4.3 Lexical biasing

The standard view argues that decisions are made in a propositional reasoning system and propagated to the motor system. If the standard story is true, then we should expect the motor representation to fulfill the content of the intention. This implies a specific and deterministic relationship between particular intentions and particular perceptual/motor representations. My position, which I have defended at length elsewhere (Burnston 2017a, 2017b) and will only summarize here, is that it is simply empirically false that there are linguistically/propositionally structured representations that bear a determinate relationship to specific motor representations. Instead, lexical representations relate to perceptual and motor ones in ways that are highly general and context sensitive, highly probabilistic, and highly associationist.

Take the perceptual category 'dog' and the motor category 'pet'. To the extent that one intends to 'pet the dog', there seems to be a straightforward relationship between the lexical representation and a particular act of petting. But this straightforwardness is highly misleading. 'Dog' corresponds to any number of subtypes of animals, with a wide range of perceptual and petting-affordance properties, and what it means to 'pet' any specific one will depend highly on the particulars of those properties. The standard story is committed to a way in which the tokening of a particular propositional attitude fixes what the perceptual-motor system will do, because those

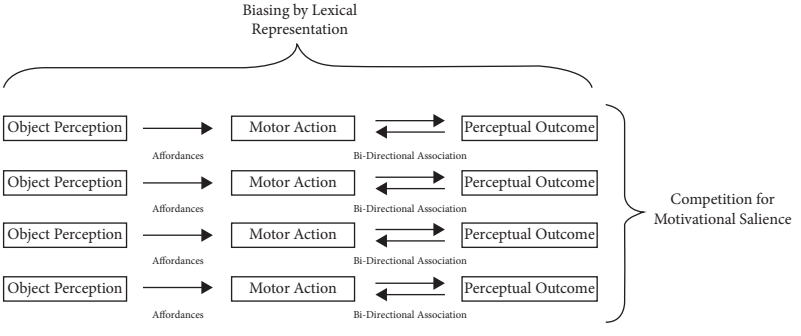


**Figure 6.3** The biased competition model

systems do not have the resources themselves to contribute to the decision. On the embodied approach, however, we can see another possible role for this relationship, namely as one that *biases* the competition between action alternatives along relevant behavioral dimensions, and a wide range of psychological evidence supports this kind of view (Burnston 2017a, 2017b; Memelink and Hommel 2013). The full picture is given in Figure 6.3.

### 6.4.4 Scaling up the embodied view

The following objection naturally arises: the kind of framework I have been discussing is based on simple object-action associations, which are triggered by the agent's immediate environment. Hence, they cannot account for *deliberative*, *future-directed*, *flexible* behavior. It is non-trivial, then, to show how the embodied approach 'scales up' to sophisticated decision-making behavior. There are three key concepts for doing so: *rich perception*, *imagery*, and *sequence representation*.

Recall that it is the perception of the environment that forms the basis of affordance competition. If all perception can represent is simple conjunctions of basic perceptual features, then there is limited basis for this kind of perception—an event file corresponding to the action 'make oatmeal' or 'obtain pizza' can't guide action unless those categories can be represented in perception. Fortunately, a huge amount of philosophical and empirical research is suggesting that perception can represent categorical and social information in addition to more basic properties (Goldstone and Hendrickson 2010; Siegel 2006; Toribio 2015). It is well established that recognizing types of objects primes motor activity associated with those objects (Tucker and Ellis 2001), and Wu (2008) has argued that this kind of enriched perception underlies our practical felicity with objects. Moreover, preparing a *type* of action (say, a grasp), primes recognition of types of objects that afford the action (Fagioli et al. 2007). So, rich perceptual-motor associations can underlie relatively sophisticated behaviors.

In the event-files methodology, it is perceived *outcomes* that trigger the associated actions. An important part of accounting for *deliberation* within the embodied framework is to propose that potential action outcomes can be represented in imagery, and thus undergo competition for motivational salience. So, when considering what to have for breakfast, I may entertain perceptual images of oatmeal or eggs. Since I 'know'—in this framework, 'have an event file for'—the actions that bring about these two outcomes,

selecting one outcome as the desirable one will trigger the relevant actions. That is, I'll perform a grasping act towards the pantry door (where the oatmeal is) rather than the refrigerator door (where the eggs are).

But scaling this perspective up requires that *sequences* of actions can be represented. Think of the process of ending up at an oatmeal breakfast as a kind of trajectory—I can't open the oatmeal before I open the pantry, and I won't succeed with the oatmeal project if I don't put the oatmeal in the bowl *before* putting it in the microwave. In the philosophical literature, Briscoe (2018) and Nanay (2016) have argued that sequences of imagery can be used in flexible behavior, and both psychological and neural evidence bears this out. First, primates are skilled sequence learners. They can learn, not just individual associations between perceptual and motor stimuli, but multiple related ones ordered in time (Catmur et al. 2009). Second, both humans and other primates are skilled at learning to recognize conditional dependencies, in which what response is appropriate to a given stimulus can depend on a *prior* cue (Koechlin et al. 2003).

These sequential associations can be primed subconsciously. For instance, Mattler (2003) had subjects listen to sounds produced either by a marimba or a piano. In one task set, subjects would have to respond to the pitch (high or low) of the sound regardless of instrument, whereas in the second they had to respond to timbre, and thus to whether the sound was produced by the piano or the marimba. Perceptual cues were associated with each task. However, in some instances, prior to the cue, another cue was presented in a masked, and therefore not consciously observable, manner. Subjects were slower and less accurate on tasks where the prime was *incompatible* with the explicit task cue, suggesting that even a subliminally presented cue influences which task set one adopts (cf., Reuss et al. 2011).

The ability to represent distal outcomes through imagery, to represent nested conditional dependencies between stimulus-response associations, and for perceptual cues to trigger abstract task types underlies, I suggest, an obvious corollary of *planning* (cf., Hommel 2013). And if this is the case, then an amodal decision maker is not required to underlie the more distal, flexible aspects of action control. There is also evidence that these principles apply even to *consumer* behavior. Watson et al. (2014) had subjects learn to associate particular perceptual cues with particular foods. Normally, eaters show what is known as *item-specific-satiety* effects. So, eating a lot of ice cream may make you sated on ice cream, even if you might still willingly consume some other food. What the researchers showed is that presentation of an

associated perceptual cue can *overcome* item-specific satiety effects. That is, the presence of an associated cue makes it more likely that subjects will choose to consume a food for which, without the cue, they would exhibit satiety. Watson et al. explicitly tie this to consumer behavior such as choosing to eat fast-foods, which is obviously a deliberative decision requiring sophisticated behavior (driving to the drive-thru, ordering the burger, etc.). As such, the kinds of principles here are potentially relevant for thinking about real-world, sophisticated decision-making.

Lastly, the framework can potentially explain even broadly 'economic' or value-driven decision-making. Krajbich and Rangel (2011) had subjects choose between three different foods in a series of forced choice tasks. Prior to the study, they had subjects subjectively rank many different potential options, and used those rankings in their competition model to *bias attention* towards certain options. Recall that input from the environment drives the competition. Therefore, attention can bias the competition process by focusing the mechanism on a certain kind of input (cf., Wu 2011). Krajbich and Rangel used the model to predict eye movements in the choice tasks. It turns out that, rather than selecting one option discretely, subjects' eyes continue to move across *all of the options* until right before the final choice, with the *proportion* of fixations to given options slowly modifying to favor the eventual choice. Krajbich and Rangel claim that this process reflects the bias that subjective value provides to the inputs, driving the competition in one direction rather than another. Intriguingly, an analogue of Bratman's notion of *commitment* is present in the models as well. Bronfman et al. (2015) showed that crossing a decision threshold leads to *decreased sensitivity* to new information. So, once a choice has been reached, the decision system stops integrating new evidence as to what will be best. In recent research, it has been shown that such models can explain even failures of economic rationality, such as preference reversals (Trueblood et al. 2014; Tsetsos et al. 2015). These results suggest that, even for conscious value-based decisions, competition models can explain human action.

## 6.5 Pluralism

Should the philosophy of action take an implementation-specific or an implementation-neutral perspective on the attitudes? In this section, I suggest that there is no easy route between our intuitive conception of the attitudes

and our cognitive architecture. The upshot, I suggest, is that we should embrace a kind of pluralism about action explanation.

There is no doubt that the standard view is the most well-worked out interpretation of the CTA, in terms of a feasible psychological architecture. So, another way to put the question about perspective: should the potential falsity of the standard view undermine attitude-explanation, or should it not? Assuming for the moment that the standard view is indeed false, there are two potential options. First, one could commit to a different operationalization of the attitudes that is not in conflict with the prevailing data. Second, one could simply embrace the implementation-neutral perspective.

Both possibilities are in tension with the embodied view, particularly the fact that our intuitive notion of 'intention' does not isolate a *single* kind or type of cause, but instead a range of possible mental states and functional roles. While I think that the perspective offered by the embodied approach explains the *fundamental* mechanisms of decision, it would be a mistake to say that only event files can count as intentions. It seems to me undeniable that, when we discuss our actions, we phrase our reasoning in terms of propositional attitudes and syllogisms. It just turns out that the closest psychological corollaries of those states—lexically structured, reportable propositional representations—are not the determiners of action. This doesn't mean they play *no* role, however. It is certainly the case that human action involves a range of linguistic devices, from instructions to mnemonics, that help structure our behavior. While I have argued that one can capture this function with the 'biasing' notion of influence between lexical and sensorimotor representations, that in no way impugns the importance of the lexical devices. Indeed, human action and skill learning would be hard to explain without them (for further discussion, see Burnston 2020).

So, the notion of 'intention' arguably does not pick out a single kind of mental state with a particular functional role. Why not simply embrace the implementation-neutral approach, then, and admit that there may be distinct kinds underlying the same notion? The real problem here is explanatory. When we explain an action by an appeal to intention, what have we explained? If we admit that the notion ranges over distinct mental kinds, then citing the general class underdetermines how the action comes about. If one rejects both of these possibilities, one might be tempted to sign on for a non-causalist view (see, e.g., Dancy 2003; McLaughlin 2013; Russell 2017). While I think non-causalist views have a lot to be said for them, in picking out the many distinct roles that folk-psychological explanation plays for us,

I also think that explaining how agency fits into the world simply must involve describing the causal processes that lead to action.

I suggest that the best approach is a pluralistic one. Attitude explanations can play a variety of roles, some of which cite causes and some of which don't. The ones that do cite mental causes range over a variety of distinct mental states and functional roles. Explanations invoking attitudes can describe more basic, non-deliberative processes, as well as complex imagery with event files, *and* explicit lexical reasoning about action. It would only be a mistake to say that these are all the same kind of mental state.

While this sounds anodyne, I don't think it leaves the philosopher of action off of the psychological hook. An articulation of intention that does not separate distinct functional roles, and attribute to each the right kind of mental state, will simply fail to describe the causal underpinnings of action. As one can probably tell, I think there are rich possibilities for understanding action and agency within the embodied approach. And while this has a reductionistic flavor, even non-reductive causalists have to admit that there must be *some* distinct realizers underlying the different causal roles they describe (Sehon 2013). The pluralist view simply demands that, when discussing attitudes, we clarify which of the potential explanatory roles, and which distinct kinds of mental causes, we are talking about.

## 6.6 Agency

There are a variety of objections one could raise for the embodied account from the perspective of the philosophy of action. The first is that, at best, the embodied account can only explain a minimal kind of agency, and leaves distinctively human agency out of the picture. The second is that the embodied account is susceptible to a particularly harsh version of the 'disappearing agent' problem (Schlosser 2011)—the worry that a causal theory of action eliminates agency by reducing it to mere causal happenings. I address each in this section.

### 6.6.1 Higher-order objections

According to some, distinctively human agency is due to *higher-order* mental states. Frankfurt (1988) famously suggests that genuine agency depends on *endorsement*. On this view, an agent is one who can reflect upon the

desires that motivate their action, and form *second-order* desires regarding the desires they wish to guide their action. Bratman ties the notion of intention to the notion of a *policy*. Agents' intentions are part of policies of self-governance, which keep the agent moving towards goals as part of evolving plans. Slightly more generally, Korsgaard (2009) argues that human agency involves a distinctive kind of reflective *self-determination*, which is absent in other organisms.

Nothing I have said rules out views like this as accounting for part of agency. One could suggest that the embodied account can only explain decision-making based on first-order states, and that it leaves these philosophical views of human agency untouched. I have two things to say in response.

First, what is the in-principle argument that the embodied approach can only explain actions based on first-order mental states? We have seen that the approach has the resources to potentially account for decisions that involve (i) temporally extended plans with nested subgoals, and (ii) value-based choices. It is an open theoretical question how much of human action can be accounted for within this perspective, but these results show that the framework at least encroaches on the kinds of actions posited by higher-order theorists. (Indeed, Watson 1975 points out that it is possible to view higher-order states as simply 'more competitors'; this fits well with the notion of competition we've been discussing.) If the limits of the embodied view are stretched, then the purportedly distinct form of agency has less and less to do. And this is important, because Bratman and Frankfurt treat higher-order states as 'primary' in defining agency. Hieronymi (2009) illustrates the point I am making here in the suggestion that the *vast majority* of human action is pursued without guidance from specific, articulated policies.

Second, one must decide whether higher-order theories are based on the implementation-neutral or the implementation-specific perspective. If the former, then again it is not clear why the embodied account cannot explain these aspects of agency—according to the implementation-neutral perspective, it is an open empirical question how agency is to be explained. If the latter, however, then the higher-order theorist takes on a significant explanatory burden, namely articulating the notion of higher-order states that is at work, and saying how it fits into the causal structure underlying decision. And, if my arguments against the standard implementational story are on track, then one option—defining higher-order states in terms of propositional structure and place in a representational hierarchy—is off the table.

### 6.6.2 The disappearing agent

I have offered a view on which the primary mechanism of decision is a causal one, of whose operations we are (at least largely) unaware, and which is subject to influences outside of our knowledge, occasionally leading to irrational behavior (e.g., failures of economic rationality). The CTA is predicated, however, on our reasons and motivations—the ones for which we can be held accountable—driving our behavior. Without this, one might think, there is no real agency to be had here. Similarly, higher-order views would object to my identification of agency with a competition between 'pulls' in the environment, as discussed in the last subsection. I reply to these concerns as follows.

First, it would be a mistake to assume that, since we are not aware of the *nature* of our decision-making mechanism, we are never aware of its operation. When we consider options, gather information, and finally plump for an outcome, we often *are* aware of the factors that are driving our decision-making system. If I decide to have a beer because it *sounds good* to do so, and I have no strong reason not to (e.g., no strongly competing work obligations), I am aware of the most salient motivation in my landscape of nearby action-outcomes. I can then, indeed, answer for the operation of this mechanism by citing the lack of competing obligations. Nothing about this requires that the neural mechanisms underlying the decision are either introspectively accessible or under my direct conscious control.

Moreover, I think that the CTA usually, and quite significantly, oversimplifies action phenomenology. In cases of *difficult* decision, or of *rushed* decisions, there is, I submit, a phenomenology that corresponds to the thresholded competition process of decision-making. When you really don't know whether you want tea or coffee, and the line behind you begins to lengthen and grumble, this increases the *urgency* of your decision (interpreted as lowering the decision-threshold; see Cisek et al. 2009), so that you plump for the option currently in mind. Now, you may further *justify* your decision by citing further reasons (the desire to lower your caffeine intake, the poor night's sleep you had yesterday, etc.). But this does not mean that, in the moment, a given set of propositional attitudes directly caused the decision.

Still, one might worry that a view that places our decisions at the whim of motivations which are cued by the environment does not allow for a notion of rational control that substantiates agency. However, there is a notion of control, which is already present in the literature, that can begin to take on

some of this explanatory work. Kennett (2001), for instance, discusses 'diachronic' control, which involves *setting up one's circumstances* in advance. This kind of explanation has already seeped into the public consciousness—the best way to avoid eating sweets, we're told, is to not have them in the house. And the best way to do that is to not go to the grocery store when you're hungry. The best way to avoid procrastinating is to remove distractions. Recovering addicts should avoid not just drugs, but the *contexts* in which they are likely to be tempted to consume (Robinson and Berridge 2008). This makes sense on the embodied view. One sets up one's circumstances so as to avoid encountering an object or context that will be motivationally salient at the time of decision. Deciding to undertake the manipulation of one's circumstances is itself, of course, a decision, but there is no bar to this kind of planning within the embodied account.

While Kennett still situates this view in a broadly hierarchical picture, Hieronymi (2009) goes further, saying that what we do in this kind of control is 'manipulate our brains', by means of altering our attention, or by persuading ourselves to 'see things differently' (p. 152). I suggest that we should take these claims literally. We supply ourselves with attentional cues, mnemonics, and imagery that will focus us on some action outcomes rather than others, predisposing our decision mechanisms to come to certain conclusions. So, when tempted by the coffee one might repeat to oneself 'only tea after 5', and thereby bias attention towards one option rather than another. (In this sense, I have analogized discursively structured intentions to a grocery list; Burnston 2017a.) When considering making a cutting remark, one might image the likely outcome of their partner's hurt expression. When considering the gym membership, one might envision one's new waistline, etc.

High-functioning agents will have mastered a variety of such strategies so that, in normal circumstances, and with exceptions, their actions flow seamlessly along with their stated values. Again, nothing I've said is incompatible with the embodied approach to decision-making. And, if the empirical picture is correct, this might be the best kind of control we can attain.

## 6.7 Conclusion

The standard approach to the CTA within philosophy of psychology seeks to situate all of the explanatory power of attitudes like intention at a particular causal nexus—namely, the transition from a practical reasoning system to a motor system that executes its dictates. I have given empirical and

philosophical reasons to question this approach. Instead, I have suggested that we adopt a pluralism about common-sense psychological notions, and their mapping to the mechanisms that produce our behavior. Doing so opens up insights about agency that will, piecemeal, validate some of our common-sense intuitions and philosophical distinctions, and vitiate others. This, I think, is the best we can hope for in situating philosophical views amongst the sciences.

## References

Bratman, M. E. (2000). Reflection, planning, and temporally extended agency. *The Philosophical Review*, 109(1), 35–61.

Bratman, M. E. (2001). *XV—Two Problems about Human Agency*. Paper presented at the Proceedings of the Aristotelian Society.

Briscoe, R. (2018). Superimposed mental imagery: On the uses of make-perceive. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198717881.001.0001/oso-9780198717881-chapter-8.

Bronfman, Z. Z., Brezis, N., Moran, R., Tsetsos, K., Donner, T. and Usher, M. (2015). Decisions reduce sensitivity to subsequent information. *Proceedings of the Royal Society B, Biological Sciences*, 282(1810), https://doi.org/10.1098/rspb.2015.0228.

Burnston, D. C. (2017a). Interface problems in the explanation of action. *Philosophical Explorations*, 20(2), 242–58.

Burnston, D. C. (2017b). Cognitive penetration and the cognition–perception interface. *Synthese*, 194(9), 3645–68.

Burnston, D. C. (2020). Anti-intellectualism for the learning and employment of skill. *Review of Philosophy and Psychology*, https://doi.org/10.1007/s13164-020-00506-5.

Butterfill, S. A. and Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, 88(1), 119–45.

Carruthers, P. (2004). Practical reasoning in a modular mind. *Mind & Language*, 19(3), 259–78.

Catmur, C., Walsh, V., and Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2369.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2), 67–90.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*(5), 801–814.

Cisek, P. and Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33, 269–98.

Cisek, P. and Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), https://doi.org/10.1098/rstb.2013.0479.

Cisek, P., Puskas, G. A., and El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, 29(37), 11560–71.

Cushman, F., Kumar, V., and Railton, P. (2017). Moral learning: psychological and philosophical perspectives. *Cognition*, 167, 1–10, doi:10.1016/j.cognition.2017.06.008.

Dancy, J. (2003). Precis of practical reality. *Philosophy and Phenomenological Research*, 67(2), 423–8.

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, 60(23), 685–700, doi:10.2307/2023177.

Fagioli, S., Hommel, B., and Schubotz, R. I. (2007). Intentional control of attention: action planning primes action-related stimulus dimensions. *Psychological Research*, 71(1), 22–9.

Ferretti, G. and Caiani, S. Z. (2018). Solving the interface problem without translation: the same format thesis. *Pacific Philosophical Quarterly*, 100(1), 301–33.

Fodor, J. A. (1983). *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Fourneret, P. and Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36(11), 1133–40.

Frankfurt, H. G. (1988). Freedom of the will and the concept of a person. In *What Is a Person?* (pp. 127–44). Dordrecht: Springer.

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66–74.

Gallivan, J. P., Stewart, B. M., Baugh, L. A., Wolpert, D. M., and Flanagan, J. R. (2017). Rapid automatic motor encoding of competing reach options. *Cell Reports*, 18(7), 1619–26.

Goldstone, R. L. and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1), 69–78.

Grafton, S. T. and de C. Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, 26(4), 590–616.

Hieronymi, P. (2009). Two kinds of agency. *Mental Actions*, 2009, 138–62.

Hommel, B. (1998). Event files: evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, 5(1–2), 183–216.

Hommel, B. (2013). Ideomotor action control: on the perceptual grounding of voluntary actions and agents. In W. Prinz, M. Beisert, and A. Herwig (eds.), *Action Science: Foundations of an Emerging Discipline* (pp. 113–36). Cambridge, MA: MIT Press.

James, W. (2013). *The Principles of Psychology* (Vol. 1). Redditch: Read Books Ltd.

Kennerley, S. W., Behrens, T. E., and Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, 14(12), 1581–9.

Kennett, J. (2001). *Agency and Responsibility: A Common-Sense Moral Psychology*. Oxford: Clarendon Press.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648), 1181–5.

Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Krajbich, I. and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13852–7.

Lavender, T. and Hommel, B. (2007). Affect and action: towards an event-coding account. *Cognition and Emotion*, 21(6), 1270–96.

Loh, M. and Deco, G. (2005). Cognitive flexibility and decision-making in a model of conditional visuomotor associations. *European Journal of Neuroscience*, 22(11), 2927–36.

McLaughlin, B. (2013). Why rationalization is not a species of causal explanation. In A. Laitinen, C. Sandis, and G. D'Oro (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*. Basingstoke: Palgrave Macmillan.

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78–84.

Mattler, U. (2003). Priming of mental operations by masked stimuli. *Perception & Psychophysics*, 65(2), 167–87.

Memelink, J. and Hommel, B. (2013). Intentional weighting: a basic principle in cognitive control. *Psychological Research*, 77(3), 249–59.

Mylopoulos, M. and Pacherie, E. (2017). Intentions and motor representations: the interface challenge. *Review of Philosophy and Psychology*, 8(2), 317–36.

Mylopoulos, M. and Pacherie, E. (2018). Intentions: the dynamic hierarchical model revisited. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1481.

Nanay, B. (2016). The role of imagination in decision-making. *Mind & Language*, 31(1), 127–43.

Pacherie, E. (2000). The content of intentions. *Mind & Language*, 15(4), 400–32.

Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, 107(1), 179–217.

Railton, P. (2012). That obscure object, desire. *Proceedings and Addresses of the American Philosophical Association*, 86(2), 22–46.

Reuss, H., Kiesel, A., Kunde, W., and Hommel, B. (2011). Unconscious activation of task sets. *Consciousness and Cognition*, 20(3), 556–67.

Robinson, T. E. and Berridge, K. C. (2008). Review. The incentive sensitization theory of addiction: some current issues. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 363(1507), 3137–46.

Russell, D. (2017). Intention as action under development: why intention is not a mental state. *Canadian Journal of Philosophy*, 48(5), 742–61, doi:10.1080/00455091.2017.1414524.

Schlosser, M. E. (2011). Agency, ownership, and the standard theory. In J. H. Aguilar, A. A. Buckareff, and K. Frankish (eds.), *New Waves in Philosophy of Action* (pp. 13–31). Dordrecht: Springer.

Schurger, A. and Uithol, S. (2015). Nowhere and everywhere: the causal origin of voluntary action. *Review of Philosophy and Psychology*, 6(4), 761–78.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Sehon, S. (2013). The causal theory of action and commonsense psychology. In A. Laitinen, C. Sandis, and G. D'Oro (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*. Basingstoke: Palgrave Macmillan.

Siegel, S. (2006). Which properties are represented in perception? In T. S. Gendler and J. Hawthorne (eds.), *Perceptual Experience* (pp. 481–503). New York: Oxford University Press.

Sims, A. and Missal, M. (2019). Perceptual decision-making and beyond: intention as mental imagery. In B. Feltz, M. Missal, and A. C. Sims (eds.), *Free Will, Causality, and Neuroscience* (pp. 13–34). Leiden: Brill Rodopi.

Smith, M. (2012). Four objections to the standard story of action (and four replies). *Philosophical Issues*, 22(1), 387–401.

Stoet, G. and Hommel, B. (1999). Action planning and the temporal binding of response codes. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1625–40.

Toribio, J. (2015). Visual experience: rich but impenetrable. *Synthese*, 195, 3389–406.

Trueblood, J. S., Brown, S. D., and Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, 121(2), 179–205.

Tsetsos, K., Chater, N., and Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: comment on Trueblood, Brown, and Heathcote (2014). *Psychological Review*, 122(4), 838–47.

Tucker, M. and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8(6), 769–800.

Uithol, S., Burnston, D. C., and Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia*, 56, 129–39.

Uithol, S., van Rooij, I., Bekkering, H., and Haselager, P. (2012). Hierarchies in action and motor control. *Journal of Cognitive Neuroscience*, 24(5), 1077–86.

Waszak, F., Hommel, B., and Allport, A. (2003). Task-switching and long-term priming: role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology*, 46(4), 361–413.

Watson, G. (1975). Free agency. *The Journal of Philosophy*, 72(8), 205–20.

Watson, P., Wiers, R., Hommel, B., and De Wit, S. (2014). Working for food you don't desire: cues interfere with goal-directed food-seeking. *Appetite*, 79, 139–48.

Wu, W. (2008). Visual attention, conceptual content, and doing it right. *Mind*, 117(468), 1003–33.

Wu, W. (2011). Confronting many-many problems: attention and agentive control. *Noûs*, 45(1), 50–76.