CHAPTER 1

Perceptual Decision-Making and Beyond: Intention as Mental Imagery

Andrew Sims and Marcus Missal

The standard view in the philosophy of action is the Causal Theory of Action (CTA). On this view, a behavioural item counts as an intentional action if and only if it is caused by the appropriate sorts of mental states in the right kind of way. On most popular contemporary accounts (e.g. Searle 1983; Bratman 1987; Mele 1992; Pacherie 2008), the appropriate sort of mental state is an intention, with the intention construed as an attitude towards a proposition. The “right kind of way” is thought to be one in which the content of the intention propagates from an abstract level of description (e.g. the intention to investigate a noise) through to more fine-grained specifications that give a bodily movement a rational structure at the moment of bodily movement (e.g. the intention to switch on this light), and finally culminating in motor commands required to execute the right bodily movements. These levels of abstraction respectively correspond to so-called distal intention, proximal intention, and motor intention. These three kinds of intention have distinct roles in the overall dynamics of intentional action (Pacherie 2008; Mele this volume).

In our contribution to this volume we offer an alternative theory of intention, on which it is not a propositional attitude at all but rather a distinct kind of mental imagery. For the purpose of our argument, we can provisionally define a mental image as a quasi-perceptual representation that occurs in the absence of the corresponding stimuli. Such imagery need not be conscious; it can also be unconscious. It may manifest in one or more perceptual modalities (Nanay 2017). The main difference that we wish to mark is that mental imagery has a quasi-perceptual format rather than a quasi-linguistic or propositional one. That idea will be developed in more detail in Sections 1 and 2.

Our account is inspired by work in the perceptual decision making literature, where decision is modelled as a process of evidence accumulation under conditions of uncertainty and noise. In the paradigms that are central to this

---

1 This work is supported by the Action Recherche Concerté project “Causality and Free Will.”

We would like to thank Anna Drozdzewska and Daniel Burnston for some helpful comments on an earlier draft.
literature, the subject is required to make a simple movement (e.g. look left or right) depending on a perceptual condition (e.g. a visual signal which indicates the “correct” direction). As time passes, the accumulation of perceptual evidence biases the drift of a single value called a decision variable, which when it hits a bound (i.e., becomes equal to a specified value, or set of values, if the variable is multidimensional) propagates a signal to the relevant effectors and causes the bodily movement. Models of this kind have great predictive power when it comes to the way that error rates (“wrong” decisions) and response times are distributed in experimental samples.

It could be argued that these studies lack ecological verisimilitude: not many decisions are of this nature (though some are, like crossing the street when the light turns green). However, it’s nonetheless possible that the model we describe could profitably be extended to contexts beyond perceptual decision making. For example, it’s been suggested that in cases where an action is more or less endogenous (i.e. not simply produced in response to a cue), the same model could be used without there being any specific perceptual evidence to be integrated with the decision variable (Schurger et al. 2012; Murakami 2014; Schurger et al. 2016). In this case, Schurger et al. (2016, 78) argue, “the process of integration to bound is dominated by ongoing stochastic fluctuations in neural activity that influence the precise moment at which the decision threshold is reached.” In other words, in these cases it is just neuronal noise which biases most strongly the decision variable, without any real perceptual evidence to speak of.

In this context, we should certainly want to know what it means for something to be intentional: if the evolution of the decision variable during spontaneous voluntary movement is dominated by stochastic drift, then this would appear to make the outcomes of our decisions implausibly random. With this in mind, our suggestion is that intentions can be understood as mental images that play the role usually reserved in this model for bona fide perceptual evidence (e.g. the onset of a visual cue). Intentional action is therefore to be

---

2 We do not mean to attribute such a view to Schurger and colleagues. Those authors do suggest that there can be biasing influence from “anticipation, subjective value [and] clock monitoring” and that “parameters, such as the movement type, may be fixed by a prior decision, which may in turn channel the neural activity leading up to the decision threshold towards a specific effector.” (Schurger et al. 2016, 78) Presumably what they have in mind when discussing neuronal noise is endogenous voluntary movements where the agent has no strong commitment to any outcome, like in Libet-style experiments. But proposals beyond these are largely unspecific at present; it’s our ambition to put some meat on these bones in a way that addresses extant problems in the philosophy of action. We are much inspired in this by Burnston’s (2017) programmatic remarks.
construed in terms of this evidence-accumulation mechanism, but with its scope expanded so that inner states as well as evidence can play a biasing role. We see the following merits in this account. First, it sidesteps some pressing problems in the philosophy of action; these are problems that we suspect result from the theory of intentions as propositional attitudes. For example, there is an issue concerning the interface between propositionally-formatted representations and the representational format that is appropriate to fine motor control. That is to say that there is confusion in the literature over how the content of a proximal intention could propagate to a motor intention (Butterfill & Sinagaglia 2014; Mylopoulos & Pacherie 2017). Indeed, some authors express scepticism that this is possible at all (Burnston 2017). On our account, there is no need to explain such an interface because the representational format that is appropriate to sensorimotor control is common to all levels of intention.

Second, it expands on the Schurger et al. (2012) interpretation of Libet-style experiments and the sceptical arguments that issue from them. The orthodox interpretation is that the readiness potential (RP) represents the final stages of action planning and preparation, and that this occurs prior to the conscious intention to move. That is why it is used in an argument that decisions are never caused by conscious intentions: the putative decision (the RP) occurs before the conscious intention. Schurger and his colleagues argue convincingly that this interpretation is wrong: we should interpret the RP as the stochastic drift of the decision variable, which ramps up to hit the bound at about the same time as the agent feels the “conscious urge to move.” Our account will offer a framework for extending this proposal to action more generally, and for the non-sceptical interpretation of Libet-style paradigms and others like them.

However, to explicate it persuasively will be challenging. There will be three philosophical objections to our thesis that we must address. First, the mechanism we describe entails that action-producing mental imagery stands in competition with bona fide perceptual evidence issuing from the senses. That obliges us to explain how action ever gets off the ground: how is it that motor imagery detailing the aimed for state of affairs can bias the decision variable in a way that moves it away from what is currently perceived?

Second, it is unclear how mental images could play the quasi-inferential role in the rational control of action that propositional attitudes plausibly can. Practical reasoning can be expressed in a deductive form quite easily if the states involved are propositional, since their syntactic structure allows this. In this way, we can understand how standing intentions could constrain practical reasoning and the implementation of other intentions. A corresponding account needs to be given for mental imagery.
Finally, we need to be able to show that perceptual or quasi-perceptual content can represent a state of affairs at the level of abstraction that is appropriate to propositions, and without themselves being propositional. For instance, when I intend to go on a holiday at some point in the future, but without any specific place in mind, I have a distal intention which specifies an action at a very high level of abstraction. But we usually think of mental imagery (like perception) as being determinate with respect to all its details, and not abstract in such a way.

These important challenges for our account will be met in the course of the paper. But we must begin by discussing the hypothesis in more detail.

1 Perceptual Decision Making and Diffusion-to-Bound Models

A neural decision variable could be defined as a compound signal containing sensory evidence, priors, value, and other quantities needed to make a categorical choice. This choice results in a rule applied to the decision variable i.e. if enough evidence is accumulated in a certain context, I will choose to do this rather than that. Decisions based on perceptual evidence are often conceptualized using diffusion-to-bound models. In these models, a hypothetical neural decision signal accumulates until it reaches a threshold value (i.e., the bound) causing movement initiation. Response time and accuracy (or the percentage of correct responses) are modelled using a noisy evidence accumulation process like a continuous random walk or diffusion (Ratcliff & Hacker 1981). The diffusion process encodes information about the experienced stimulus and other sources of information present in the context that are needed to make a decision (Figure 1.1).

This theory was initially developed to model experimental results in a sequential letter-matching task where human subjects have to judge whether two strings of letters are identical. Typically, the two strings of letters are rapidly presented sequentially on a computer screen facing the subject. The subject has to compare the previously encoded string of letters held in working memory with the last one. The decision the subject has to make is: “Are the two strings the same or different?” Technically, this type of task is often referred to as “two-choice serial reaction time task” (2-CSRTT).

According to Ratcliff and Hacker (1981), if the diffusion process reached the upper bound, a “same” response was produced. Alternatively, if the diffusion process hit the lower bound a “different” response was produced. The diffusion rate towards the upper bound increases as a function of the proximity of the two strings and increases towards the lower bound as a function of the
Figure 1.1 A (top-left): This is an illustration of the visual stimulus that is presented to the subject during the dot motion paradigm. During a single trial the dots all move in random oblique direction except for a specified proportion which move either to the left or to the right. In the trial shown, 10% of the dots (those which are darkened) are moving to the right; B (top-right): This shows the evidence that the visual stimulus of Fig. 1.1A is taken to comprise. It is a normal distribution with mean = 0.1 and standard deviation = 1. The sensorimotor parameter is the direction of movement of the dots in the visual field. In this simulated and merely demonstrative example, the relationship between the stimulus and the distribution should not be taken as necessarily accurate; C (bottom): This shows the activity of the diffusion-to-bound mechanism over a single trial. At each time-step, a random value is sampled from the distribution in Fig. 1.1B and added to the decision variable. This continues until the variable hits the bound corresponding either to left-hand movement or right-hand movement. In this case, the decision variable hits the move-right bound after about 80 time-steps.
dissimilarity of the two strings. Moreover, human subjects probably implicitly adjusted the position of the boundaries and the zero point of the diffusion process (bias) based on previous experience or expectation. For instance, giving an instruction to the subject before data collection (e.g. answer “same” only if you are sure versus answer “different” only if you are sure) changes the bias in the model and therefore reaction time is prolonged but accuracy is increased.

Several experiments in the Rhesus monkey have tried to find areas in the brain where the accumulation of evidence and/or the decision variable could be processed. The major problem is to differentiate accumulation of sensory evidence from a decision signal at the single neurone level. Indeed, the two signals are inevitably linked. Therefore, perceptual decision making paradigms have been developed in order to separate in time these two different signals (see the review in Gold and Shadlen 2007). For instance, in the random-dot motion paradigm, Rhesus monkeys or human observers have to decide whether a set of random dots move to the left or to the right. The percentage of coherently moving dots is randomly and parametrically varied between successive presentations (Gold, J.I., Shadlen, M.N. 2002). Even at zero coherence (all dots moving randomly) subjects must make a decision.

Shadlen and co-workers have shown that a neural correlate of the decision variable involved in this process can be found in the parietal cortex (area LIP; Shadlen et al. 2006; Roitman & Shadlen 2002). Interestingly, neuronal activity also reflects the choice of subjects at zero coherence (no sensory evidence per se). Activity in area LIP is compatible with predictions of the diffusion-to-bound model. Lafuente & Romo (2005) have reached similar conclusions using a task in the somatosensory domain where Rhesus monkeys have to detect a 20-Hz sinusoidal indentation applied on the finger pad. This simple detection task was accompanied by neuronal recording either in the primary somatosensory cortex (S1) or the medial premotor cortex (MPC). S1 neurons encode the presence of the sensory stimulus, but not the decision variable. On the other hand, MPC neurons are good candidates to encode the decision of the monkey in the presence (or absence) of the vibrotactile stimulus.

All these experimental paradigms provide strong electrophysiological support for the diffusion-to-bound theory of decision making. However, most real life decisions involve deliberation over topics where repeated experience is usually not available (e.g., a first marriage). Perceptual decision experiments are often rather simple, linking one particular sensory quantity to one movement (e.g. a button press, an eye movement). It is therefore tempting to posit different mechanisms for decision making, one for simple sensorimotor decisions and one for more complex deliberate decisions (Kahneman 2002).
However, we suggest that it is more plausible to postulate one decision making mechanism that could be instantiated in different areas depending on the context. Indeed, in humans even simple decisions are actually often influenced by deliberation. The aim of the present paper is to speculatively extrapolate from the theoretical contribution of diffusion-to-bound models to the kinds of decisions discussed by philosophers of action.

2 Extending the Model

As noted, perceptual decision-making paradigms do not accurately represent all decisions made by human beings in the wild (i.e. outside the laboratory), except perhaps in a highly idealised way. That is because human beings in the wild also perform actions more or less endogenously, that is to say, without any obvious perceptual cue that directly triggers the decision. Such choices usually also take place over more than two alternatives. However, the models for perceptual decision making can be extended to account for these more complex cases. Such extension is relatively speculative but it is not without support.

In the paradigms discussed above, there are usually just two alternatives. For example, in the dot movement paradigm the choice is whether to indicate that the dots are moving left, or whether to indicate that they are moving right. One of these choices will be incorrect, depending on the perceptual cue (e.g. if I indicate left when the dots are moving right). Simple choices like these are easy to model: as we have shown, this can be done in two dimensional space where one decision corresponds to time and the other corresponds to the value of the decision variable. In such a two dimensional model we can specify two bounds that correspond to two outcomes, but we cannot add any more.

Therefore, to model more complex decisions requires that we add more dimensions to our model. That allows us to add more bounds in order to represent more possible outcomes. For example, we might model a decision variable that drifts stochastically through three dimensional space. Strictly speaking, the decision variable will have two dimensions <x, y> while the third dimension <z> represents the evolution of the decision variable through time. In this three dimensional space we can add more reachable bounds that represent more possible outcomes. The bound will be defined by a hyperplane, which in a three dimensional space is just a subspace with two dimensions. At each step in the drift of the decision variable, two values are added (plus the time step) that dictate how far it moves in this space. Theoretically speaking there is no limit to the number of dimensions we can add to our model, despite the fact
that it becomes difficult to visualise after the fourth is added. So in principle, we can represent any number of possible outcomes within this diffusion-to-bound style of model.

The dimensions of our model correspond to sensorimotor parameters that vary continuously along a scale. Examples of these parameters from vision may be the size of an object, or its location in the visual field. Similarly, these parameters may encode for motor features that have continuous properties, like extension of reach or strength of grasp. For a particular bodily movement, then, the dimensions which are integrated into the decision variable at each moment reflect the sensorimotor parameters for which these dimensions encode. This allows us to improve on our provisional definition of “mental image.” In the context of this paper, we propose to understand a mental image as a set of values in \( n \) dimensions which correspond to a range of sensorimotor parameters, where this range is minimally limited to a specific value in each dimension that the image includes, but which may also specify a wider range. So, for example, a mental image may specify a set of highly constrained values that correspond to my extending a finger to touch a specific point in front of me, but it may also specify a looser range of values that correspond to my having an arm extended out in front of me, without any particular point being specified.

Hommel’s (2004) theory of event files suggests a specific way in which we might construe the role of sensorimotor evidence during endogenous action. Event files are episodic bindings of perceptual features, a task context, and motor features. Event files specify a particular directed bodily movement. For example, it might specify the action of reaching for a bottle of water. In this case, the event file would bind the perceptual features corresponding to the action (proprioceptive, visual, and other sensory items associated with successful reaching) with motor commands that correspond to the correct extension, rotation, and grasp.

In the context of diffusion-to-bound models, it may be that we can understand the bounds as corresponding to event files, the sense that the event files specify the appropriate locations in state space for the execution of a particular action. Then, different influences bias a decision variable in the state space by accumulating evidence for one or another event file. The evolution of the decision variable through the state space represents the competition between event files for implementation in behaviour (cf. Cisek & Kalaska 2010). The outcome of that competition is jointly determined by stochastic drift (i.e. neuronal noise) and biasing influences. These influences are understood here to include the following two sorts of item: i) sensorimotor values that are caused by stimuli (i.e. perception per se); and ii) sensorimotor values that are not
caused by stimuli (i.e. mental images). These two sorts of items both serve as biasing influences that drive the decision variable to a bound.

Here is a simple case which is intended to make the proposal clearer. You are standing in front of a door. Then you open the door and walk through. What happened? A certain kind of cta-theorist will say that you had an intention with propositional content corresponding to something like “open the door and walk through.” This intention propagated from higher cognitive areas into motor cortex and was somehow translated into the appropriate motor commands, and then bodily movements.

Our story is different. The right bodily movements correspond to a set of sensorimotor values that are executed in the right sequence. The decision variable does not correspond to these parameters before you act; it is rather fluctuating around the values associated with the current perceptual state (the closed door). Action begins when biasing influences begin to push the variable towards the values associated with the successful execution of the action. The influence in question will be a set of endogenously generated quasi-perceptual representations—that is, a mental image—that is implemented in probability distributions over the appropriate values in those sensorimotor parameters. The decision variable is continuously updated with randomly sampled values from these distributions, and this drives the value of the decision variable to the event file that corresponds to you having opened the door.

A story like this can be extended upwards to more paradigmatically intentional action and it can also be extended downwards to what philosophers call “sub-intentional” action (O'Shaughnessy 1980). Sub-intentional action is purposive activity that is not intentional under a description, like idly doodling on a notepad while focused on something else, or tapping impatiently while waiting for something. Actions of these kinds can be construed in our model as biasing influence that results from learned association of perceptual values with corresponding “stereotyped” motor values. For example, it may be that a pen lying next to a pad is associated with the action of doodling, and so automatically potentiates the sensorimotor parameters associated with doodling. This is what should normally happen in the absence of motor inhibition (cf. Duque et al. 2017). In cases where we are focused on something else, we fail to inhibit those actions that are potentiated. Furthermore, the capacity for such motor inhibition may be reduced by brain lesion, in cases such as in utilisation behaviour where the affected part of the body engages with nearby objects in a way that is contextually inappropriate (Shallice et al. 1989).

The upward extension is more complex. It requires that endogenous sensorimotor variables can bias the evolution of the decision variable at great temporal and spatial distances from the corresponding action. For example, if I want
to walk to the cafeteria on the other side of campus to get myself a coffee, then this requires that I am able to potentiate the right sensorimotor parameters in the right sequence, and in a way that respects contextual demands (e.g. not taking a route that I know to be closed). We think that in these more complex cases propositional attitudes have a role to play, but it is not the role that is envisioned by CTA-theorists. We propose to elucidate this upward extension of our account by addressing a few key objections from the perspective of CTA.

3 Some Conceptual Problems with the Account

First, it’s not clear why mental imagery would sometimes win competition with percepts that are immediate to the agent. Secondly, we might doubt that mental imagery as a representational format is of sufficient conceptual complexity to be able to exercise the rational control that is required of an intention. Finally, despite its evidential support in simple contexts, we might doubt that models for perceptual decision making could be expanded to contexts in which we intend to perform an action that is specified at a very high level of abstraction. In this section of the chapter we address these concerns.

3.1 Can Mental Images Produce Action?

A critic of our view might deny that mental imagery can drive action, which would make it unsuitable for playing the role of intention. The reasons that might be cited for this view are as follows. First, there is the commonsense idea of mental imagery and imagination in general as a kind of derivative version of ordinary perception. The similarity between imagery and perception is noticed by many authors in the history of philosophy, most notably by Hume (1739) in his claim that impressions and ideas are distinct only in their “force and vivacity,” and indeed something like this comparison motivates representational views of perception in which the representational content is made to do work in distinguishing hallucination from veridical perception.

Secondly, it might be argued that the default attitude to hold upon the apprehension of some content in imagination—insofar as any attitude is taken up at all—would be belief, rather than intention or desire. That is to say, if I imagine something, and (in some exceptional case) I forget or I am unable to know that I am merely imaging, then surely it is more correct to say that I believe my eyes, rather than that I intend my eyes. So here, the objection will be to say that for the mental image to play the role of an intention requires something more than the image. And presumably the suggestion will be that the additional ingredient is some sort of conative intention-like attitude towards that image, which would make our account redundant.
However, these objections and others like them rely on a commonsense notion of mental imagery or imagination which we are not committed to. On the mechanism we have hypothesised in Section 2, what is given as input to the mechanism is just a set of sensorimotor parameters which correspond to the consequences of the action, and which drive it to a bound. This means that there is no real representational distinction of type between a perceptual representation and the endogenous representation (“intention”) that produces action. This feature of our approach—despite how odd it may seem to philosophers—is not unprecedented. Prinz (1997), for example, outlines the “common coding hypothesis,” on which perception and action planning share a representational form in common. That is to say that items that are “efferent” (i.e., motor) and those that are “afferent” (i.e., sensory) are continuous and overlap; this would mean that none of these representations code specifically for afference or for efference, or if they do, then they code for both simultaneously, as in Millikan’s (1995) hypothesised “pushmi-pullyu” representations.3

There are empirical reasons to think that this common coding hypothesis is true. If “event codes” and “action codes” encoded respectively for afference and efference, then we would expect their respective functional roles to be well defined. Accordingly, we would not expect there to be interference or facilitation effects on action by way of perception, or vice versa. The interference hypothesis states that when a particular code (e.g., corresponding to location) is in use for perception, it will be impaired in its use for action (and vice-versa). The facilitation hypothesis states that dimensional overlap between a stimulus and a response can enhance the speed of the response. These sorts of effects are often reported in the literature (see Prinz 1997; Hommel 2004 for reviews; cf. Burnston 2017), which motivates for the view that direction of fit is not part of these representations.

All this, of course, obliges us to answer the following question: when, and in virtue of what cause, is it endogenous mental imagery rather than perception which brings about action? That we have a mental image in working memory must certainly not be sufficient for bringing about action; sooner or later we must accept that such images simply do not correspond with the deliverances of perception. So how do such images can drive action, in the absence of an encoded direction of fit that allows it to be distinguished from perception? Someone taking our view seriously might think that veridical perception prevents action ever getting started, because this constant source of evidence will bias the decision variable in preference to any endogenous imagery.

3 This corresponds to the philosophical distinction between “directions of fit” (see Anscombe 1957, §32 for a classic exposition).
We can develop the mechanism of Section 2 in order to answer this question. The inputs to the decision variable at some time consist of an array of values that correspond to sensorimotor parameters. These values, however, are not discrete: given the conditions of uncertainty and noise that prevail, these values are more likely to be probability distributions. Probability distributions have various properties. One of these properties is precision. The precision of a probability distribution is inversely related to its standard deviation; it is a measure of uncertainty. A distribution which represents a perceptual parameter with a very high level of uncertainty will have a low precision; values that are sampled from that distribution will not cluster very tightly around the mean. The converse is true of distributions with a high precision (Figure 1.2).

With this in mind, our suggestion is that whether or not some mental image is perceptual or action-driving is a function of the precision of the corresponding

---

4 We are inspired by similar ideas in the context of predictive processing (e.g., Hohwy 2013; Clark 2016). However, nothing we say here commits us to that framework as a whole.
distributions. Specifically, mental imagery will produce action, rather than be revised on the basis of sensory information, in cases where it is higher in precision than the sensory information that is in tension with it. When it has a high precision the values that are integrated into the decision variable will be closer to the goal-state value than if the distribution is imprecise. Similarly, when the current perceptual state is relatively imprecise then the values that are sampled from that state will be relatively ineffectual in preventing the biased diffusion of the decision variable to the goal state.

To conclude this section of our argument, we accept that our initial account was problematic in the context of philosophical orthodoxy regarding direction of fit. However, we believe that this distinction is only important in the present context because it explains why a content would sometimes be afferent (responsive to informational change) and sometimes be efferent (action-producing). In response to this concern, we have suggested how we can explain this in the context of an account that does not rely on the idea that there is an action/perception distinction built into the representations that underlie cognitive processes.

3.2 Content and Rational Control
The second conceptual problem that we wish to address corresponds to the role of intentions in allowing rational control over action as it unfolds in a particular situation. This control has a number of different aspects. Pacherie (2008), following Buekens et al. (2001), classifies these into two. The first is tracking control. Tracking control is our ability to rationally control an action as it unfolds in time. For some action, it will include various steps whose temporal sequence must be respected, and it may be that we are required to implement an intention in an unforeseen manner if we are thwarted at any of these steps. Take the action of fetching a cup of coffee. To successfully perform this action I have to leave my office, get to the kitchen, and set up the coffee machine to make a cup. Tracking control of this action will involve performing this action in the right sequence (e.g., not trying to extract the coffee before the ground beans are in place) and improvising on the fly (e.g., if there is work being done in the corridor, I will have to take a detour to the kitchen from my office).

The second is collateral control. This involves controlling for side effects of my action. When I go to fetch my cup of coffee I may wish to perform the action in a way that respects other commitments that are in the background. For example, it may be that I implement the intention by taking a circuitous route in order to avoid encountering someone I don’t wish to see. I may also wish to implement my intention in a way that respects etiquette: if there is someone
already waiting for the coffee to brew, then I will not cut in line to take my cof-
fee before they do. The bottom line for both of these types of control is that
they need to be responsive to, i) other commitments I have, even if disposi-
tional (the commitment to acting politely); and ii) the exigencies of the pres-
ent context in the here-and-now.

For certain instances corresponding to both of these sorts of control, it may
be argued that the intention needs to be poised to play a role in practical rea-
soning. If I find myself thwarted by contextual constraints (work in the corri-
dor), then I will need to engage in practical reasoning on the basis of the
intention in order to revise my plan of how best to implement it. For collateral
control, my intention needs to be rationally responsive to more abstract disposi-
tional states regarding certain constraints within which I would like to bind
my behaviour. So it would appear to be the case that both of these types of
control involve inferential relations between the intentions and propositional
attitudes like beliefs and desires.

Given the requisite inferential sophistication of intentions, one criticism of
our proposal might be that mental images are inapt to play that role. That is
because inference seems to be something that is appropriately performed in
an amodal conceptual format, rather than a quasi-perceptual format. It is hard
to see how operations like material inference, disjunction, and so on could be
implemented in the latter, but these are arguably required in order to exercise
rational control over action. To see this, say that I have the intention to get a
cup of coffee and there are two different paths that I could take to the kitchen.
I also want to pass by my colleague’s office on the way, and this is on one of the
paths. So I somehow have to be able to implement the intention in a way
which respects this other intention, and that requires that my intention is ra-
tionally responsive to the other. If intentions are conceptual or language-like,
then it is easy to see how this would work: just in the same way that inference
is performed in written or spoken language. But this is not possible on our
account.

Our reply to this objection is to advance a clarification. It is not a conse-
quence of our position that propositional attitudes can play no role in action
production whatsoever; our contention is just that they do not play any causal
role *qua* intentions. It may well be the case that propositional attitudes play a
key role in relations between images, and this would not necessitate that inten-
tions are propositional attitudes. For example, van Leeuwen (2013) argues that
propositional attitudes like beliefs can inferentially govern transitions between
imaginings. Here’s how he demonstrates the point:

Instructions: imagine a female lion in the bushes looking out on a herd of
grazing antelope; then imagine she charges them. What do you imagine
next? I’m guessing you imagine the antelope run away. But why? Imagining this wasn’t in the instructions. You could have imagined them staying put, though you didn’t. The answer, around which some consensus has emerged, is that one’s beliefs inferentially govern transitions from initial imaginings to later imaginings.

Van Leeuwen 2013, 226

We can imagine a wide range of examples of this basic type, where the way mental imagery unfolds in time is governed by our dispositional mental states. I can “deliberate” in imagination when it comes to settling on an intention (Nanay 2016), and my deliberations will be constrained by what I believe it is possible for me to do, and what I think is the likely result of some imagined action. Furthermore, we see no reason to limit this governing role to beliefs. It seems that desires may play a key role in imaginary transitions, as well. Certainly, this role is familiar enough from the quite widespread and ordinary experience of pleasant daydreams. In daydreams, transitions between imaginings are often governed by what we would like to happen, and also by our beliefs about potential ways in which such things might occur.

The constraints that propositional attitudes place on transitions in imagery provide us with a way to understand how mental images qua intentions could exercise the kinds of control discussed above: tracking control and collateral control. Let’s take the previous case of going to get a cup of coffee. As I implement the intention, it’s necessary that I implement it in a way that respects a certain sort of temporal sequence. If I try to extract the coffee before I load the ground beans into the machine, then my action will not succeed. This is rational control as tracking control. Mental images can exert this kind of control since transitions between images is governed by our beliefs about the world. I have certain beliefs about the way this coffee machine works, and these beliefs govern the way in which my imagery-intention represents the successive stages of the action that I am engaged in.

Likewise, when I intend to get my cup of coffee, I need to implement my intention in a way that respects my other commitments. If I go and get it in a way that violates these commitments, then I will not be exercising collateral control over my action. Collateral control can be explained by the way that our desires govern transitions between mental images. For example, I may have the background commitment to talk to a colleague on my way to the kitchen. This is to say that I desire that I talk to my colleague. This desire will govern the way my intention-imagery represents successive stages on an action, by governing the transitions between image stages. So the intention will be that I go by the circuitous route to the kitchen on which my colleague’s office is located.
In short, there is no reason to think that images cannot exert the right kinds of rational control over action once the role of bona fide propositional attitudes is taken into account. Our claim is just that propositional attitudes do not themselves play the role of intentions, as the prevailing orthodoxy states. Why then, would this not make our account of intentions as mental images superfluous? In other words, aren’t we back now at an orthodox view that includes both mental images and propositional attitudes? That is not the case; our position differs in a subtle but important respect. Specifically, on our account there is no propagation of content from propositional attitudes to mental images (i.e. motor representations), whereas that is a central feature of CTA (cf. Pacherie 2008). For us, propositional attitudes merely play an auxiliary role in constraining possible transitions between mental images.

The reason that this difference is important is because a key puzzle that drives contemporary philosophy of action—the “interface problem”—concerns how it is that content could propagate from propositional representations to sensorimotor representations. Pacherie’s framework is subject to this problem, because for her there is such a propagation. But for us there is not, and we are therefore providing a way of dispensing with this vexing philosophical problem about content.

A critic might still object that we are committed to some propagation of content as soon as we accept the view that propositional attitudes could govern transitions between images. That is because such governance requires that the content of the proposition plays some sort of causal role in the transition, and this amounts to the propagation we claim to be entitled to deny. To this objection we can give a suggestion about how this might occur in the absence of direct translation from propositional content to sensorimotor content.

Let’s return to Van Leeuwen’s example of the lion and gazelles. In this case, there is a transition between the image of the lion charging these gazelles to an image of the gazelles fleeing the lion. Let’s say that this transition is governed by a belief that gazelles (generally) flee from lions. Our critic will say that what happens here must be some sort of propagation of content between that relevant proposition and the image, because the proposition seems to have causal influence over the evolution of the episode of imagining. And that would oblige us to give an answer to the interface problem.

However, we don’t see that mere causal influence amounts to content-propagation. For example, it may be that the proposition <gazelles flee from lions> activates a number of high level sensorimotor parameters without there being any direct translation of the propositional content into sensorimotor imagery. For example, it may be that due to associative processes the tokening of the lexical item which expresses that proposition biases the decision variable
towards the range of sensorimotor parameters which correspond to a statistical regularity that holds between its contents (c.f. Burnston 2017): i.e., that a large animal charging at a group of smaller animals is usually followed by the scattering of the latter in flight. Furthermore, it’s important to note that on our account the intentions are sensorimotor images all the way up and down the hierarchy of intention, and propositional attitudes only play a role in delimiting the transitions which may occur between these images. In Dretske’s (2010) terminology, propositions are “structuring” causes and not “triggering” causes.

One may object nonetheless that this is a version of the causal theory after all, since propositional attitudes are exerting some causal influence over decisions. To this we believe the correct response is just to draw attentions to the details of our account. What we are interested in establishing is that intentions can be understood as mental images (not propositional attitudes), and that the mechanisms for bringing about action should not be understood on the model of the practical syllogism. If that is nonetheless to be labelled a causal theory of action, then in our view it is so distinct from the orthodox causal theory in its details as to make the shared label misleading. The cta as ordinarily pitched is committed to much more than the bare thesis that propositional attitudes play some sort of causal role in action.

3.3 Can Mental Imagery Be Abstract?

Intentions can specify actions at a very high level of abstraction. Someone can have an intention to succeed in their career, to find true love, or to make the world a better place. Such a level of abstraction seems to require propositional content. That is because an intention which is so abstract is compatible with very many sorts of imagery. What would the imagery be that corresponds to the intention to make the world a better place? That is an abstract idea that does not seem to correspond to any specific image whatsoever. However, it is ideally represented in a propositional format, and we can easily see how an intention could propagate this abstract format into more concrete and intentions proximal to action if it were conceptual in form.

However, it is a mistake to assume that mental imagery cannot be abstract. Consider Block’s (1983) discussion of the “photographic fallacy” in debates over mental imagery. One objection posed to theorists of mental imagery is based upon the idea that imagery-style representations must be determinate with respect to all their details. For example, Pylyshyn (1978) discusses the fact that when four-year olds are shown and then asked to draw a tilted beaker containing coloured water, they do not draw the water as remaining horizontal but rather draw it as perpendicular to the beaker (as if it did not respect gravity). He claims that this result motivates for the idea that visual experience is
“descriptive” (by which he means quasi-linguistic) rather than “pictorial,” since if the visual representation in memory were pictorial then it should specify visual details determinately, in this case, the angle between the surface of the liquid and the beaker.

In response to this line of argument, Block replies that there is no need that pictorial representations must be determinate with respect to every visual feature that they represent. Sometimes, they can represent the existence of some determinable property (e.g., the way that mountainous terrain is represented on a map) while being non-committal on the specifics of the property (e.g., the number of mountains, and their specific topography). What sets the image apart from a proposition or conceptual content is rather that it is iconic in its format. It lacks syntactic structure of the kind that language possesses. To make this difference stark, note that there is a syntactically correct way to decompose sentences. It does not do to split the sentence “Bob is happy” right down the middle, since it will make nonsense of the parts “Bob i” and “s happy.” By contrast, an iconic representation (like a map) can be decomposed in any of number of ways without respect for any sort of syntax, and the parts will continue to represent just as well as the whole (Fodor 2007; cf. Quilty-Dunn 2016).

With this in mind, we are happy to put ourselves on the line in claiming that there can be intentions representing actions at a very high level of abstraction. We make sense of this in terms of a hierarchy of parameters, and for which specifications of values for the higher parameters will automatically activate a range of values at the lower levels in the hierarchy, but will leave them more or less unspecified within this range. For example, one may have a mental image of a drinking glass that is indeterminate with respect to the precise geometric dimensions of the glass. However, there will be a number of dimensions that will be primed on account of their likelihood relative to others. If prompted, we can make these dimensions more determinate, and presumably some similar process is at work for unconscious mental images, as well. We contend that the making determinate of lower level parameters corresponds to the means-end reasoning that is characteristic of practical reason once an agent is faced with implementing an abstract intention in a specific situational context.

It may be difficult to see how a picture would work like this in specific cases, like that of the intention to “make the world a better place.” We think the right characterisation of these will need to proceed on a case by case basis. In this

---

Note that this sets our account apart from that of Mylopoulos and Pacherie (2017); although they also talk in terms of determinate and determinable properties, they take the representation of determinable properties by intentions to be propositional while we take them to be iconic or quasi-perceptual.
particular case, the content seems to be so vague that very many things come under the description. In fact, we would suggest that even on a propositional view of intention it will be difficult to explain what the content of this intention is. On our view, it is likely that there is no particular sensorimotor content that is appropriate to this content, but it also seems that if the agent were to deliberate on the meaning of this sentence then it would eventually result in a more concrete set of plans which could constitute a bona fide image-as-intention, on our view.

In any case, our main aim in this subsection is to dispute the idea that mental images cannot be abstract per se. That is a prejudice, since mental images can specify an indeterminate set of parameters or higher order variables that range over multiple parameters without ceasing to be quasi-perceptual in format. We can see the transition from highly abstract (or “distal”) intentions to “proximal” intentions implementable in a concrete situation in terms of the progressive filling out of the lower-level parameters in a hierarchy, until such time as the image is of an appropriate grain to drive the decision variable to bound in the diffusion to bound mechanisms discussed in Section 2.

4 Conclusion

Diffusion-to-bound models of decision making are able to accurately model performance and response time in a wide range of perceptual decision-making paradigms. There is also very strong electrophysiological support of a model of this general kind. It is therefore a good place to begin in understanding the neural mechanisms which underlie decision-making in general.

Some theorists will demur from this assessment, on the grounds that intentional action has a character wholly different to the kinds of tasks in experimental paradigms modelled using diffusion-to-bound. In the dot motion paradigm, for example, the subject is required to make a decision about whether the dots are moving left or right. But in the cases that are paradigmatic in the philosophical literature, an agent is required to choose between two more courses of action on the basis of competing reasons. In some cases, for instance, the choice is between a course of action that is pragmatically best and one that is morally best (e.g., the choice to help someone at some cost to oneself). So one may refuse the relevance of diffusion-to-bound models on the basis that they seem to lack the rational component of choice that philosophers want to understand.

However, this ruling of diffusion-to-bound mechanisms as non-rational in principle seems to us premature. There are good reasons why the extrapolation
of diffusion-to-bound modelling ought to be attempted in spite of the superficial differences between intentional and perceptual decision-making. Primarily, to posit mechanisms distinct from those that are well evidenced in simpler contexts is to be committing oneself to a dual-process account on which different mechanisms are implemented for different sorts of decision making (Evans & Stanovich 2013). This is not an invalid path to take, but it comes with theoretical costs.

Primarily, positing multiple mechanisms obliges us to explain how those mechanisms interact in order to produce behaviour. With respect to intentional action, this explanation will need to be twofold. Firstly, it will need to explain how it is that one or another mechanism for action is selected over the other(s) that are posited by the account, and in what circumstances; secondly, it will need to explain how the representational format that is putatively appropriate to higher-level intention interfaces with the representational format that is putatively appropriate to bodily movement (Burnston 2017). Our account sidesteps these issues.

A second cost to this approach is that it risks indulging in exceptionalism about human action that is at odds with the view that we occupy a non-exceptional place in the natural world. Although it is true that human behaviour appears to have a remarkably wide scope and flexibility, it seems more plausible to think that this is the result of progressive iterations on simpler mechanisms that we share with other animals, rather than a discontinuous leap from those mechanisms. These costs, while clearly not fatal to the dual-mechanisms approach or to cta, seem to us to justify the removal of objections against alternative approaches.

References

Anscombe, G.E.M. (1957). Intention. Oxford: Blackwell.
Block. N. (1983). The photographic fallacy in the debate about mental imagery. Nous, 17, 651–661.
Bratman, M. (1987). Intentions, Plans, and Practical Reason. Cambridge: Harvard University Press.
Buekens, F., Maesen, K., & Vanmechelen, X. (2001). Indexicaliteit en dynamische inten-
ties. Algemeen Nederlands Tijdschrift voor Wijsbegeerte, 93, 165–180.
Burnston, D.C. (2017). Interface problems in the explanation of action. Philosophical Explorations, 20, 242–258.
Butterfill, S.A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. Philosophy and Phenomenological Research, 88, 119–145.
Cisek, P., & Kalaska, J.F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33*, 269–298.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Davidson, D. (1963). Actions, reasons, and causes. Reprinted in *Essays on Actions and Events*, 1980. Oxford: Clarendon Press.

De Lafuente, V., & Romo, R. (2005). Neuronal correlates of subjective sensory experience. *Nature Neuroscience, 8*, 1698–1703.

Dretske, F. (2010). Triggering and structuring causes. In T. O’Connor & C. Sandis (eds.), *A Companion to the Philosophy of Action*, ch. 18. Oxford: Wiley-Blackwell.

Duque, J., Greenhouse, I., Labruna, L., & Ivry, R.B. (2017). Physiological markers of motor inhibition during human behaviour. *Trends in Neurosciences, 40*, 219–236.

Evans, J.S., & Stanovich, K.E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science, 8*, 223–241.

Fodor, J.A. (2007). The revenge of the given. In B.P. McLaughlin & J.D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, pp. 105–116. Malden: Blackwell.

Frith, C.D., Blakemore, S.-J., & Wolpert, D.M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London B*, 355, 1771–1788.

Gold, J.I., & Shadlen, M.N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron, 36*, 299–308.

Gold, J.I., & Shadlen, M.N. (2003). The influence of behavioural context on the representation of a perceptual decision in developing oculomotor commands. *Journal of Neuroscience, 26*, 632–651.

Gold, J.I., & Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Hommel, B. (2004). Event files: feature binding in and across action and perception. *Trends in Cognitive Sciences, 8*, 494–500.

Hume, D. (1739). *A Treatise of Human Nature*. Oxford: Clarendon Press, 1978.

Kahneman, D. (2002). Nobel Prize lecture: maps of bounded rationality, a perspective on intuitive judgment and choice. In T. Frangsmyr (ed.), *Nobel Prizes 2002: Nobel Prizes, Presentations, Biographies, & Lectures*, pp. 416–499. Stockholm: Almqvist & Wiksell Int.

Mele, A.R. (1992). *Springs of Action: Understanding Intentional Behaviour*. Oxford: Oxford University Press.

Millikan, R.G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives, 9*, 185–200.

Murakami, M., Vincente, M.I., Costa, G.M., Mainen, Z.F. (2014). Neural antecedents of self-initiated actions in secondary motor cortex. *Nat. Neurosci.*, 17, 1574–82.
Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: the interface challenge. *Review of Philosophy and Psychology*, 8, 317–336.

Nanay, B. (2016). The role of imagination in decision-making. *Mind & Language*, 31, 127–143.

Nanay, B. (2017). Multimodal mental imagery. *Cortex*, http://dx.doi.org/10.1016/j.cortex.2017.07.006.

O’Shaughnessy, B. (1980). *The Will: A Dual-Aspect Theory*, vol. 2. Cambridge: Cambridge University Press.

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

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129–154.

Pylyshyn, Z. W. (1978). Imagery and artificial intelligence. From C.W. Savage, ed., *Perception*.

Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies*, 23, 255–263.

Ratcliff, R., & Hacker, M.J. (1981). Speed and accuracy of same and different responses in perceptual matching. *Perception and Psychophysics*, 30, 303–307.

Roitman, J.D., & Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22, 9475–9489.

Schurger, A., Mylopoulos, M., & Rosenthal, D. (2016). Neural antecedents of spontaneous voluntary movement: a new perspective. *Trends in Cognitive Sciences*, 20, 77–79.

Schurger, A., Sitt, J.D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *PNAS*, 42, E2904–E2913.

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

Shadlen, M.N., Hanks, T.D., Churchland, A.K., Kiani, R., & Yang, T. (2006). The speed and accuracy of a simple perceptual decision: a mathematical primer. In K. Doya et al. (eds.), *Bayesian Brain: Probabilistic Approaches to Neural Coding*, pp. 209–237. Cambridge: MIT Press.

Shallice, T., Burgess, P.W., Schon, F., & Baxter, D.M. (1989). The origins of utilisation behaviour. *Brain*, 112, 1587–1598.

Van Leeuwen, N. (2013). The meanings of “imagine” part 1: constructive imagination. *Philosophy Compass*, 8, 220–230.