

Unconscious Perception and Central Coordinating Agency

Joshua Shepherd and Myrto Mylopoulos

Abstract: One necessary condition on any adequate account of perception is clarity regarding whether unconscious perception exists. The issue is complicated, and the debate is growing in both philosophy and science. In this paper we consider the case for unconscious perception, offering three primary achievements. First, we offer a discussion of the underspecified notion of central coordinating agency, a notion that is critical for arguments that purportedly perceptual states are not attributable to the individual, and thus not genuinely perceptual. We develop an explication of what it is for a representational state to be available to central coordinating agency for guidance of behavior. Second, drawing on this explication, we place a more careful understanding of the attributability of a state to the individual in the context of a range of empirical work on vision-for-action, saccades, and skilled typing. The results place pressure on the skeptic about unconscious perception. Third, reflecting upon broader philosophical themes running through debates about unconscious perception, we highlight how our discussion places pressure on the view that perception is a manifest kind, rather than a natural kind. In doing so, we resist the tempting complaint that the debate about unconscious perception is merely verbal.

1 Introduction

By broad consensus across cognitive psychology, neuroscience, and philosophy, unconscious perceptual states exist. They are posited to explain the results of a wide range of studies (e.g., those involving masked priming), and to help explain otherwise puzzling phenomena (e.g., blindsight). And the existence of unconscious perceptual states is widely taken to have important implications for the nature of perception (Burge 2010, Berger and Nanay 2016, Phillips 2018a), for the epistemic role of consciousness in perception (Siegel 2016, Berger et al. 2018; Smithies 2019), for how we understand the functions of perception (Peters et al. 2017), and for how we understand the functions of consciousness (Rosenthal 2008).

Consensus has recently come under sustained attack (see Peters & Lau 2015; Peters, Ro, & Lau 2016; Peters et al. 2017; Phillips 2018a). According to the ‘problem of the criterion’, cases of seemingly unconscious perception are actually cases of weakly conscious perception that goes

unreported by subjects simply because they are not confident enough in what they consciously perceive.

This problem is not easy to overcome. A natural move is to restrict oneself to objective measures of consciousness, such that a subject is considered consciously aware of a stimulus if they perform above chance at a discrimination task pertaining to that stimulus, and not consciously aware of it otherwise. The idea is that if we can demonstrate the existence of sensory representation of some stimulus through downstream effects on behaviour (e.g., priming effects), even though a subject's discriminatory capacity is completely at chance, then we can infer that the representation at issue is unconscious. There would be no need to rely on verbal report. The problem of the criterion would be avoided.

But here a second problem arises. This is what Phillips (2018a, 21) calls the 'problem of attribution' (see also Phillips 2016, Phillips & Block 2016, Peters et al. 2017). In many – arguably, in all – of sensory representations influencing behavior while discriminatory capacity is at chance, the representations do not qualify as genuine *perceptions*. They are but sub-personal states of, e.g., the visual system.

This articulation of skepticism about unconscious perception has generated a flurry of recent activity (see, e.g., Quilty-Dunn 2019; Phillips 2019; Phillips 2020; Berger & Mylopoulos 2019, forthcoming; Phillips forthcoming), as philosophers, psychologists and neuroscientists reconsider issues once thought put well to bed. The discussion is already rich, but has in our view given insufficient attention to assumptions about agency that underlie the problem of attribution. Possibly as a result, the discussion is over-reliant on findings from vision science. Our aim in this paper is to re-focus reflection on issues that might move understanding forward.

Here is how we will proceed. In the next section we discuss a notion that plays a key (if somewhat subterranean) role in much of the debate thus far. This is the notion of central

coordinating agency. Plausibly, a sensory state will only qualify as perceptual if it bears the right relation to central coordinating agency. But getting the nature of this relation right is non-trivial. We review options and offer an explication of the relation that, we argue, has the potential to clarify the case-interpretations that drive much of the back and forth thus far.

In section three we turn to results from vision science that Phillips charges fail, due to the problem of attribution, to provide evidence for unconscious perception. While Phillips's points are well-taken in many cases, we offer a novel interpretation of a key result that, we argue, puts pressure on Phillips. One way it does so is via a more subtle understanding of the architecture of central coordinating agency. We elucidate this more subtle understanding and argue that thinking of central coordinating agency in this way brings a wider range of research to bear more directly on the issue of unconscious perception.

We illustrate this latter claim in section four. There we review an interesting experimental paradigm on skilled typewriting, and we argue that the results arguably provide evidence of unconscious non-visual – specifically, haptic – perceptual states. One interesting idea that emerges from this discussion is that in order to find the best interpretation of the status of sensory states in skilled typing, one might profitably consider roles metacognition might play – beyond that of setting confidence thresholds – in shaping participant judgments.

In the paper's conclusion, we reflect on broader philosophical themes running through debates about unconscious perception. In particular, we highlight how our discussion places pressure on the view that perception is a manifest kind, rather than a natural kind. Finally, we consider whether the debate about unconscious perception can be considered merely verbal.

2 Central coordinating agency

The problem of attribution turns on the distinction between personal and sub-personal states and processes: states and processes that are attributable to an *individual as a whole* as opposed merely to some subsystem of that individual (see Dennett 2002 for the original distinction and Drayson 2014 for recent discussion of it). We are happy saying, for instance, that I eat the apple, but my stomach digests it, or that you understand the words that are coming out of my mouth, while your language faculty parses the syntactic properties of the utterance. Similarly, while I see the red apple, my visual system computes information about its reflectance properties and light intensity. But these intuitive cases, while motivating the thought that there is a distinction to be drawn here, do not yet guide us towards a principled basis on which to draw it.

Developing ideas found in Burge (2010), Phillips (2018a) considers the idea that attributability of a representation to an individual could be understood, roughly, in terms of its availability to *central coordinating agency*. This availability is meant to play an epistemic role insofar as it ‘comprises our best evidence for attributing a representation to an individual’ (494).

Fair enough. But what is it to say that a representation is available in this way? We get an idea of how Phillips thinks of this through his discussion of various cases.

In a discussion of Jiang et al. (2006), which shows that nudes rendered unconscious by continuous flash suppression impact reflexive spatial attention, Phillips implies that such impacts do not qualify as individually attributable: ‘stimulus-driven, reflex-like attentional responses’ do not ‘count as manifestations of central agency’ (25). Call this Idea One: reflex-like processes do not manifest central coordinating agency.

In a discussion of results like Norman et al. (2013), which demonstrate that attention can be directed to objects not consciously seen, Phillips claims that the attentional processing at issue is ‘akin to a stimulus-driven reflex, operating entirely outside of voluntary, agentic control’ (25) and that if subjects cannot use representations (of the unseen but attended objects) to guide responses,

then central coordinating agency is not implicated. Call this Idea Two: if an agent cannot use a representation or representational process R to guide behavior, then R is not available to central coordinating agency. In a more recent paper, Phillips endorses a stronger version of this idea, replacing behavior with intentional action: ‘The key (at least evidential) criterion for such individual attribution is, I propose, availability of the representation for guidance of intentional action’ (Phillips 2019, 7).

In fairness to Phillips, these ideas are offered to illuminate particular cases. Phillips is not offering an analysis, or account, of central coordinating agency. And in the cases Phillips discusses, these ideas are plausible. But we want something that could guide theorizing with more consistency.

Consider Idea One: ‘Stimulus-driven, reflex-like’ processes do not manifest central coordinating agency. One might worry that it is not terribly clear at this juncture what to include in the class of ‘stimulus-driven, reflex-like’ processes, but on any intuitive understanding of such processes, cases of them manifesting central coordinating agency exist. Skilled agents do not seek to rid themselves of all stimulus-driven, reflex-like processes. Rather, skilled agents train up reflex-like habits tuned to specific types of stimuli for specific reasons and specific situations. Reflex-like behaviours that are well-integrated with an agent’s capacities in some action domain can be very useful, and can then be put to work in flexible ways.

Phillips’s use of ‘stimulus-driven’ and ‘reflex-like’ was intended to characterize the capture of spatial attention in a context in which the capture had little to do with what the agents were otherwise thinking or doing. In that context, his claim about availability to central coordinating agency has a ring of truth. But the reason may bear little relation to the process’s being reflex-like. It may instead have to do with the relation the process bears to the control the agent exercises over her behavior in a situation.

Idea Two again: if an agent cannot use a representation or representational process R to guide behavior, then R is not available to central coordinating agency. This link between use for the purposes of guidance and availability to central coordinating agency has been suggested before. In a now-classic paper, Frankfurt (1978) puts guidance front-and-center in distinguishing between those things that an agent does and those that merely happen to them—what he calls the ‘problem of action’. For Frankfurt, a full account of action must ‘... explain the notion of guided behavior [and] specify when the guidance of behavior is attributable to an agent and not simply, as when a person’s pupils dilate, to some local process going on within the agent’s body.’ (Frankfurt 1978, 159). This is because, on his view, action just *is* movement guided by an agent: ‘the performance of an action is [...] a complex event, which is comprised by a bodily movement and by whatever state of affairs or activity constitutes the agent’s guidance of it’ (ibid).

The challenge now is to specify in what availability for such guidance consists. We have to be careful that the availability criterion is not too strong. For example, Gareth Evans maintains that some sensory representation can be attributed to an individual only if it serves ‘as the input to a thinking, concept-applying, and reasoning system’ (1982, 158). The problem with this criterion, Phillips observes, is that ‘It is plausible to think that only representations above the objective threshold are cognitively accessible in any sense’ (2018a, 23). But the defender of unconscious perception should want to avoid requiring that representations be above this threshold, since if they are we run into the problem of the criterion. This line of reasoning is what leads Phillips to consider the weaker notion of ‘available to central coordinating agency’ as paradigmatic of ‘attributable to the individual.’

We need a less demanding notion of availability for guidance. A minimalistic option links availability with coherent contributions to the exercise of control. We could say, for example, that a representational state or process R is available for guidance in circumstances C if R can be tokened

in C and if the operation of R in C is coherently sensitive to the content of the other representational states involved in the agent's exercise of control over her behavior.

Here, then, is where we are. We propose availability to central coordinating agency as a sufficient condition for attributability to the individual.

ATTRIBUTABILITY: an agent J's representation or representational process R is attributable to J if R is available to J's central coordinating agency.

We understand availability as follows.

AVAILABILITY: an agent J's representation or representational process R is available to J's central coordinating agency if J is able to exploit R for purposes of guidance.

And since this explication relies on a term that might mislead, we offer this clarification of exploitability.

EXPLOITABILITY: an agent is able to exploit some representation or process R for the purposes of guidance in circumstances C if R can be tokened in C and R is disposed to contribute to the agent's exercise of control in C in part because it is coherently sensitive to the content of the representational states involved in such control or such states are coherently sensitive to R.¹

¹ We do not commit to any specific account of the representational states involved in control, although intention and attentional states are paradigmatic here. Indeed, it is worth noting overlapping consensus between our proposal regarding agency and control and Wayne Wu's account of agency and control. According to Wu (2011, 2014, 2016) we have control when behavior is appropriately sensitive to the agent's intention, and a key mechanism for explaining behavioral sensitivity is a process whereby intention influences attention, and attention identifies

This is, as noted, only a proposed sufficient condition for attributability. There is wiggle room for those who favor a fuller, or even an alternative, account. (This is an issue to which we return in section five.) But our explication seems fruitful, since [a] it captures some of the central motivating ideas of the leading critic of unconscious perception (namely, Phillips), [b] it is not so strong as to beg the question against unconscious perception, and [c] it illuminates one way the proponent of unconscious perception could respond to Phillips. If one can find unconscious representations that meet these criteria, and are otherwise candidates for the label ‘perceptual,’ then one will have found a plausible candidate for unconscious perceptual representation.

3 Vision-for-action, Automaticity, and Attributability

One of the most promising lines of evidence for unconscious states that meet the criteria for attributability comes from studies examining the visual representations that are carried by the dorsal stream and constitute so-called ‘vision-for-action’ (cf. Block and Kentridge in Peters et al. 2017). These representations are thought to contribute to the online control of behaviour and the programming of detailed motor commands that specify fine-grained features of our movements. They are contrasted with representations that are realized by the ventral stream, which are thought to primarily contribute to the recognition and categorization of objects in one’s environment as well as the selection of appropriate action types (e.g., grasping, throwing, picking up) directed towards them. While ventral stream states are presumed to be available to conscious experience, representations of the dorsal stream are viewed as unconscious.

targets for action, in effect assisting intention in its guiding function. We need not adopt this framework here. But it is useful to note that Wu’s work is not only in broad concordance with our proposal, it also furnishes biologically plausible explication of ways that attentional and perceptual states qualify as available for guidance.

Why think that dorsal stream vision is unconscious? Two sources of empirical evidence are often invoked. The first involves visuomotor pointing tasks. In these tasks, participants are asked to reach towards a target that unexpectedly ‘jumps’ from one position to another mid-reach (Bridgeman et al. 1979; Goodale et al. 1986; Castiello et al. 1991). In a frequently cited paper from 1991 by Castiello et al. titled ‘Temporal Dissociation of Motor Responses and Subjective Awareness’, participants were asked to correct their reaching movements following the sudden displacement of the target, and to report the time at which they became aware of its displacement by saying ‘tah!’. The experimenters found that the relevant vocal utterance occurred more than 300 ms after participants had already begun adjusting their movements in response to the target jump. On the assumption that these rapid adjustments are guided by visual representations subserved by the dorsal stream—a plausible assumption given the functional role assigned to such states—it would seem that these are good candidates for perceptual states of which the agent is not aware.

A second source of evidence comes from brain lesion studies involving individuals with visual form agnosia—a condition that results from damage to IT in the ventral stream. The most commonly discussed case of this condition is that of DF, who is unable to report on features of objects in her environment, but exhibits sensitivity to such features in the course of action guidance. For instance, in a widely discussed ‘mailbox’ task (Goodale & Humphrey 1998), during which DF was asked to ‘post’ a card into a slot oriented at different angles, she exhibited preserved ability to reach towards and orient her hand appropriately in the direction of the slot, as well as preserved grip scaling ability. But she did not seem to have conscious access to the visual orientation of the slot, as evidenced by a task in which she was asked to draw it, which resulted in what looked like random guesses.

In response to these types of cases, Phillips urges that ‘the pertinent modulations of behavior (e.g., grasp aperture) do not witness genuine control and guidance by the individual, and so

fail to meet relevant conditions for perception proper' (2018a, 30-31).² Why not? Phillips maintains that his assessment is 'strongly suggested' by the types of metaphor that theorists tend to use in describing the role of the dorsal stream, which point to the automaticity of the behaviour that it guides. For instance, it has been characterized as 'an automatic pilot' (Pisella et al. 2000), 'a tele-assisted semiautonomous robot' (Goodale and Humphrey 1998: §9, Milner and Goodale 2006 [1995]: §8.2.3), and a 'heat-seeking missile' (Campbell 2002: 56). For Phillips, dorsal stream representations are thus 'confined to the autonomous robot—a subsystem of the individual' (2018a, 501).

The central inference underlying Phillips' argument here is that if some behaviour is automatic then it is not attributable to the agent as a whole, but rather to a subsystem of the agent. This move may seem attractive, since examples of automatic behaviour that come most readily to mind are often basic reflexes that may indeed be best understood in this way. But upon closer examination, we have good reason to doubt its validity. We now explain why.

First, let's take a step back and consider what it is for some behaviour or process to be automatic. On classic conceptions in psychology (see, e.g., Posner & Snyder 1975; Shiffrin & Schneider 1977), automatic behaviour or processing is sharply contrasted with that over which cognitive control, and is thought to be marked by a stable set of properties, such as: proceeding in the absence of conscious thought or deliberation; involving little to no working memory, selective attention or cognitive effort; and being fast and efficient. Over the years, however, numerous theorists have challenged this dichotomous and absolute conception of automaticity and come to recognize that (i) rather than being all-or-nothing, automaticity is best understood as a *scalar* property

² We note that another natural response to such cases is to deny the claim that dorsal stream vision is unconscious. See, e.g., Wu 2020, for a recent argument to that effect on the grounds that the claim depends on introspection, which is not a reliable source in this context. We do not here weigh in on this broader debate (for additional contributions, see Clark 2001, Mole 2013, Wu 2013a, Schenk & McIntosh, 2010, Shepherd 2016) focusing instead on Phillips' challenge from the failure of attributability, which he views as the "more promising response" (2018: 500).

that comes in degrees, (ii) the diagnostic features of automaticity together form a *cluster* of properties that tend to, but need not always, co-occur (see, e.g., Moors & De Houwer 2006; Fridland 2014; Moors 2016), and (iii) one and the same behaviour or process can be automatic *relative* to some feature or set of features, but intentionally controlled relative to some other feature or set of features (see, e.g., Wu 2013b; Wu 2015).

This more nuanced and qualified approach to automaticity makes it doubtful that we can use this notion to draw firm conclusions about the attributability of a representation or process to an individual. Especially relevant for illustrating this point is (iii) as stated above and its accompanying analysis by Wu (2013b).

Wu (2013b) rightly notes that, rather than simply viewing automaticity and agency as diametrically opposed, we must “embed automaticity within a theory of agency”. This is especially clear when we consider cases of skilled action, which characteristically involve a tight interplay between automatic processes and agentic control (see, e.g., Fridland 2017, Shepherd 2019, 2021, Mylopoulos & Pacherie 2020).

How might we do this? On Wu’s view, agency is characterized as the process by way of which an agent solves what he calls the ‘Many-Many Problem’. This is the problem an agent confronts of selecting an appropriate path through a behavioural space defined by the attention-mediated coupling of potential targets and behavioural responses in a way that guided by the agent’s intention. So described, full-fledged instances of agency are not incompatible with some aspects of the relevant behaviour being automatic. All this means is that these aspects do not themselves constitute the solution to the Many-Many Problem that is causally structured by the agent’s intention, though they may nonetheless be aspects of the very same behaviour that does. For instance, if an agent intends to reach for a target, their behaviour constituting a successful path through the behaviour space is an instance of agency with respect to its being an instance of, e.g.,

the action type of reaching for a target, and an instance of automaticity with respect to its being an instance of, e.g., reaching with a specific force, velocity, or spatial trajectory.

The foregoing suggests that the correct way to view the role of the unconscious visual states of the dorsal stream is as subserving automatic *aspects* of behaviour, especially those involved in parameter-specification, where other aspects of that behaviour are nonetheless part of the agent's solution to the Many-Many Problem given their intention and thereby attributable to them. What we have then is a case for blocking the inference from automaticity to failure of attributability that Phillips invokes.

In light of these remarks, let us reconsider the study responsible for the auto-pilot characterization of the dorsal stream. This is a study by Pisella et al. (2000), and it is frequently cited to support the idea that vision-for-action, and in particular representations in the posterior parietal cortex, functions automatically in a way that contradicts one's conscious commands. If this is correct, it would suggest that the behaviour is *fully* automatic, i.e., relative to all aspects and in such a way that it is not sensitive to the agent's intention, which would be a problem for our claim that it is a reasonable candidate for attributability.

In Pisella et al.'s study, participants were presented with a target that could unexpectedly jump to the left or to the right. The experimenters tested the participants' abilities to interrupt a pointing movement towards that target mid-reach. They were divided into two groups. The 'location-stop' group was instructed to interrupt their ongoing pointing movement in response to the target jump. By contrast, the 'location-go' group was instructed to correct their movement in response to the target jump. The findings were that a significant percentage of corrective movements occurred despite the 'stop' instruction in the location-stop group.

How should we interpret these results? Should we go along with Phillips and conclude that here dorsal stream vision appears to subserve automatic behaviour that is not sensitive to an agent's intentions and thus not guided by the agent? Here is what Pisella et al. had to say.

[I]n response to the perturbation, the number of responses classed as corrections to target 2 significantly increased within the same movement duration of 200 ms (binomial $p < 0.05$) in both location-go and location-stop groups. This similar timing suggested that the earliest corrections observed in both groups resulted from an identical visuomotor guidance, which was independent of instruction and thus automatic. (730)

But wait. Two questions should be considered. First, the percentage of stop and go participants who corrected their movements in response to target perturbation was 9 percent and 30 percent, respectively. So, if the target-following response is fully automatic in a way that is insensitive to intention, why was it so much more common in the go condition? There is no good answer. On decent explications of automatic, one wants to conclude that the response is not fully automatic in this way.

Second, why did subjects tend to make errors – to follow the target by pointing when they were supposed to stop – only when their movements were relatively quick, i.e., under 300ms? At overall movement times above 300 ms, subjects did not make errors. In the stop condition, they reliably stopped their movements. But wouldn't an auto-pilot be engaged by the visual perturbation of the target no matter this difference in movement time?

The full pattern of results suggests the following interpretation. Subjects in these conditions deploy different response strategies, i.e., intentions. Obviously there will be variance, but we can split subjects roughly into 'incautious' and 'cautious' response groups. So, in the stop condition,

most subjects may have this strategy: try to inhibit the target jump correction, and wait for a conscious representation to stabilize. A minority of subjects may instead have this strategy: reach for the target (relatively quickly, or incautiously), and update if the target jumps. This would explain why 9 percent of subjects fail to stop, and follow the target incorrectly. Notice that even within this 9 percent, if this is the strategy, the following reaction is not counter to their intention. The failure is rather one of intention-updating, which of course takes time.

In the go condition, subjects may have this strategy: reach for the target (cautiously or incautiously). Here, the unconscious representation of the target-jump will facilitate success, even in quick reaching. In the go condition, subjects were much more likely to reach towards the target in the 200-300 ms movement window than subjects in the stop condition. This suggests some ability to inhibit the response in many subjects. A good explanation of the difference in inhibition between subjects is a difference in strategy (i.e., in intention).

Is our appeal to differences in strategy ad hoc? We think not. It is not only plausible, but consistent with a growing awareness of the pervasiveness of strategy use in studies of sensorimotor adaptation and action control. As McDougale et al. note in a 2016 review, ‘Until recently, strategy use has been considered a nuisance in studies of sensorimotor adaptation, and experimental instructions are often designed to actively discourage this behavior’ (536). But this is a mistake. Strategy use is common, and very difficult to rule out. Better to adapt research to make strategy-use as explicit as possible (see Krakauer et al. 2019 for thorough discussion).

Notice, in this connection, a study by Cameron et al. (2009) which added the following condition. Participants were told to ignore the target jump and continue to reach to the target’s original location. In this condition, participants were largely successful, deviating in only 7% of cases, compared with a 36% deviation rate when the instruction was to stop in response to the target jump. Again, a difference in intention seems to be the key explanatory difference.

What might the Phillipsian make of this study in light of our interpretation? Another natural response to such cases is of course to deny the claim that dorsal stream vision is unconscious. There is much debate regarding this issue that shows no signs of abating (see, e.g., Wu 2020, for a recent argument to that effect on the grounds that the claim that dorsal stream vision is unconscious depends on introspection, which is not a reliable source in this context; for additional contributions to the debate, see Clark 2001, Mole 2013, Wu 2013a, Schenk & McIntosh, 2010, Shepherd 2016). We do not here wish to weigh in. Rather, the point we want to emphasize in this section is that Phillips' challenge from the failure of attributability, which he views as the "more promising response" (2018: 500) fails to provide compelling reason to reject unconscious dorsal stream vision.

Furthermore, in our view, the broader pattern of work on action guidance and sensorimotor adaptation in the past 20 years indicates a picture that is compatible with the interpretation of results that we have presented here.³ Action guidance is subserved by a hierarchical computational architecture that supports functional specialization and differentiation, but also continued contact, between levels. The mechanisms of central coordinating agency operate over not only the states and processes that guide patient, explicit planning of behavior, but also, to at least some degree, the states and processes that guide sensorimotor adaptation and more fine-grained motoric implementation of plans.

So, for example, studies that examine how agents update their pointing location in response to non-veridical feedback indicate that two kinds of representational process are working at once (Day et al. 2016, McDougle et al. 2016). There are representational processes that drive judgments about task success – whether the agent's pointing behavior fulfilled her general goal. And there are

³ There is too much work to cite judiciously. Nice recent reviews of the science include McDougle et al. 2016, and Krakauer et al. 2019. For some application of the science to issues in philosophy of psychology and action, see Butterfill and Sinigaglia 2014, Christensen et al. 2015, Christensen et al. 2016, Fridland 2017, Fridland 2019, Mylopoulos and Pacherie 2017, Mylopoulos and Pacherie 2019, Shepherd 2015, Shepherd 2019, Burnston 2017, Brozzo 2017, Ferretti and Zipoli Caiani 2019.

representational processes that drive sensorimotor adaptation – where in space the agent will actually point, given an intention to point to a particular location. The former representations are paradigmatically conscious, and allow a comparison between the agent’s intended goal and the result. The latter representations are arguably non-conscious, and allow a comparison between the sensory prediction made by sensorimotor adaptation processes, and the sensory result (i.e., the visual representation of the result of the pointing behavior).

These processes are dissociable in odd experimental conditions, but they are in general tuned to work together to guide behavior in broadly rational ways. Sensorimotor adaptation processes work to implement the instructions provided by higher-level planning states, and utilize sensory feedback to suit these purposes. Arguably, they do so intelligently (Fridland 2017). Certainly, it is difficult to affirm an account of central coordinating agency that carved off these processes from the story.

Now, it may be true that these sensorimotor adaptation processes – ones like the processes driving target-following in the Pisella et al. case – operate semi-autonomously. But semi-autonomy does not entail that such processes are not available to, or even partially constitutive of, central coordinating agency.

In the next section, we discuss a targeted illustration of these points about central coordinating agency. We do so because the case usefully highlights issues that deserve further consideration if we are to move towards a better understanding of the operation of central coordinating agency and the implications for any account of perception.

4 Unconscious Perception

Our argument will go as follows. First, we elucidate a series of effects surrounding sensory registration of action error. We argue that these effects indicate that the sensory states involved are

attributable to the agent. Then, we point to two kinds of experiment that seem to indicate unconscious sensory registration of error. The conclusion follows. But in considering a response, we turn to metacognition.

4.1 Sensory registration of action error is attributable to the agent

The phenomenon we have in mind is sensory registration of action error. When agents make an action error on some task, a suite of effects reliably follows. Agents tend to slow down in their execution of the next action. This is called post-error slowing (PES).⁴ At the same time, action error is associated with increased neural activity (measured by fMRI) in sensory areas relevant to the error (Danielmeier et al. 2011), as well as with signature neural responses as measured by EEG, in particular, an early error-related negativity (ERN) and a later positive deflection of the EEG signal known as the Pe. While the Pe is usually associated with conscious perception of error, the ERN is typically present whether or not participants report awareness of error (see, e.g., Endrass et al. 2007, Ficarella et al. 2019). In the period following the error, motor suppression can also be found by measuring corticospinal excitability (Amengual et al. 2013). And physiological measures suggesting an orienting response can be linked to the redirection of attention towards the salient event (in this case, the error) (Barceló et al. 2002, Ullsperger et al. 2010).

What do the effects in this suite have to do with each other? First, they all seem causally related to the sensory registration of error. This is often assumed by experimenters, but it is worth being explicit about it here. This set of measurable effects can be elicited for a wide range of actions in a wide range of psychological and physical contexts, and what is common amongst them is the commission of error, which must be registered somehow. The standard picture of how this goes is

⁴ For debate about the functional role of PES, see Dutilh et al. (2012), Ullsperger and Danielmeier (2016), Wessel and Aron (2017).

that control mechanisms compare representations of predicted events (e.g., actions) with detected events. In the case of actions, the predictions may be informed by motor commands, or copies of motor commands sent to the control mechanisms (Wolpert 1997). But the registration of error requires the comparison of sensory feedback, which comes in via sensory modalities – e.g., in some cases visual, in some cases auditory, in some cases haptic.

Now, the inherited view, at least in philosophy, is that comparisons between predicted and registered events are conducted by a ‘sub-personal’ comparator mechanism. Here, for example, is John Campbell explicating the basic idea.

It is important that this comparator mechanism is viewed as a sub personal cognitive mechanism which affects the subject's experience of agency. People do not in general know the detailed contents of the movements they are making, so the person is not, in general, in a position to make detailed assessments of the comparison between their motor instructions and the exact movements performed. That information is typically possessed and used only by relatively low-level cognitive mechanisms, themselves remote from consciousness. (2004, 478)

Interestingly, this inherited view is in tension with much work on error perception⁵, which treats conscious perception of error as the norm and unconscious registration of error as an outlier. But the important point for our purposes is that we not pre-judge whether the sensory registration of error is sub-personal or not. For that is precisely what is at issue.

⁵ It is also in tension with recent empirically informed work on skilled action, which grants a larger role to conscious experiences than previous models allowed (see Christensen et al. 2016).

This leads to our second point, which is that the best interpretation of the data is that registration of error by sensory systems leads to a coordinated cascade of broadly rational changes in behavior that support rational action guidance. These changes include inhibition of motor representation, a change in attention, and a preparedness to update intention (or task set). Why think this cascade is coordinated? First, because elements of the cascade reliably co-occur across many different types of action errors, involving different actions and parts of the body. Second, because elements of the cascade can be linked to coordinated systems with fairly well-mapped functional roles. The case for this second point has been made in recent review papers by Wessel and Aron (2017) and Wessel (2018), which propose and argue in some detail that registration of error leads to recruitment of a frontobasal ganglia system for inhibition, as well as a brain network responsible for attentional re-orientation.

That the many effects of error registration are coordinated need not indicate that they are broadly rational. The case for this latter point is largely based upon charitable reconstruction of an agent's behavior following error. The broad pattern of effects upon motor suppression and attentional re-orientation need not enhance some particular behavioral indicator – like accuracy – on the next trial. Indeed, given evidence that response inhibition competes with neural resources (Chiu and Egnér 2015), and can negatively impact working memory, time-constrained tasks that require maintenance of specific information in working memory may be negatively impacted by registration of error (see also Notebaert et al. 2009). But in order to qualify as broadly rational, we do not need some behavioral marker to be enhanced across all action-types and circumstances. Rather, the behavioral effects of error registration should reliably put the agent in a better position to succeed – a better position to satisfy important goals – across a wide range of circumstances. We submit that this is pretty clearly the case with this particular suite of effects. As Wessel has argued, ‘errors induce a quick, automatic cascade of processing, which includes a momentary inhibition of cognitive (and

motor) processing as well as attentional orienting . . . this automatic cascade promotes a shift in attention that is beneficial to (and potentially even necessary for) the initiation of subsequent, slower, adaptive, controlled post-error processes that are geared at tuning the task set to improve accuracy' (Wessel 2018, 16).

The sensory registration of error that drives the related cascade of effects, then, arguably meets our criteria for EXPLOITABILITY, and thus AVAILABILITY and ATTRIBUTABILITY. Note that two features have to be satisfied for the registration of error to qualify as exploitable. First, the registration must contribute to the agent's exercise of control. It does so by engaging processes of motor suppression, attentional orientation, and subsequent action planning. Second, the registration must contribute to control in part because it is coherently sensitive to the content of representational states involved in control. The registration is itself sensitive, of course – it is the product of a comparison of what is happening with what is predicted to happen, according to the agent's action plan. And the registration's downstream effects on motor representations, attention, and intention also display coherent sensitivity.

The question, now, is whether sensory registration of error – the sort of registration that engenders this broad suite of effects that assist rational action guidance – is ever unconscious. And the answer is 'probably so.' We make the case for this answer by looking at two different kinds of study and their results.

4.2 Sensory registration of error is sometimes unconscious

The first kind of study involves rapidly corrected action errors, often involving saccades. Endrass et al. (2007) put participants in front of a computer screen. In the task they would first see a fixation cross and two dotted square frames. The cross would disappear and the squares would remain for 200 ms. In a cue condition, one of the square's dotted lines would increase in brightness

and width. Then the original fixation display would return for 50 ms. Next a circle would appear in either the left or right square for 100 ms. In 600 trials, the cue was incongruent with the location of the circle; in 100 trials, the cue was congruent; another 100 trials had no cue. Participants were instructed to quickly and accurately antisaccade to the mirror position of the circle in the square that contained no circle. 900 ms later participants were asked whether their performance had been correct, incorrect, or whether they were unsure. Behavioral metrics were recorded using eye-tracking technology, and neural metrics were recorded using EEG.

Participants rarely reported ‘unsure’ (1.8% of trials), and rarely reported error when performing correctly (2.6% of correct responses). That is, they rarely missed a correct response. Interestingly, however, a large proportion (around 60%) of errors were not reported—that is, they missed a large proportion of errors. In 8.6% of all cases participants correctly reported their errors, while in 13% of all cases they did not.

This difference between ‘aware errors’ and ‘unaware errors’ was not statistically significant, but Endrass et al. report other reasons to think awareness of error is importantly different from unawareness of error. Participants corrected 89% of unaware errors, but only 33% of aware errors, with quick antisaccades. In the case of unaware errors, the corrective movement was 112 ms faster on average. In addition, unaware errors had lower saccade amplitude. And aware errors were associated with PES on the following trial, while unaware errors were not. And while both aware and unaware errors correlated with an early error-related negativity (ERN), a later signal – the Pe, which the experimenters analyzed in a window of 400-600 ms after the error – did differ significantly depending upon awareness.

But there is also reason to think that unaware errors are registered as genuine errors by the sensorimotor system monitoring eye movements, and that this registration initiates action guidance processes. For while unaware errors did not generate PES on the following trial, they did generate

corrective movements. And when experimenters looked at the latency of the corrective antisaccades compared with correct cases, they found that the corrective antisaccades were slower than in trials when the correct saccade was initially made. This suggests that unaware errors are cases in which a weak amplitude saccade is erroneously made, followed by a rapid correction, and accompanied as well by a standard signal of error registration, the ERN. As Endrass et al. reason, given that participants were very accurate in reporting correct responses, rarely took the ‘unsure’ option, and that the overall discrimination accuracy of participants was high ($d = 1.41$), ‘unaware errors might actually be trials in which participants had no conscious representation of their erroneous performance’ (1719).⁶

We anticipate a charge of weakly conscious representations, or of a low confidence criterion. So it is important to point out that, in general, it seems that registration of error can take place even when participants have full confidence that no error was detected. Consider a recent study in the literature on ‘partial errors’ – errors that are, in some cases, only partially committed, and that are quickly corrected for. (Endrass et al. consider the possibility that the initial erroneous saccades in their study were just such errors.) Ficarella et al. (2019) had participants perform a button-pressing task, and measured muscle activity with EMG as well as neural activity with EEG. They also asked participants to report, after each trial, whether they made the correct move, a partial error (whether they began to press the wrong button), or an outright error. These reports embedded confidence – participants reported on a scale of 1-6 where 1 indicated ‘I am sure I did not produce a partial error’ and 6 indicated ‘I am sure I did produce a partial error.’

⁶ This basic result has been found as well in Taylor and Hutton (2011). And recent work by Willeke et al. (2019) suggests that this corrective movement is an ‘express’ or rapid microsaccade – a kind of small eye movement found to occur between 60-120 ms after onset of a stimulus (see also Tian et al. 2018) – generated by implicit registration of error.

Replicating earlier work (Rochet et al. 2014), Ficarella and colleagues found that roughly 35% of the time participants confidently detected their partial errors (responded 5 or 6) and roughly 50% of the time they did not, i.e., they were confident that they did not make a partial error when they had (responded 1 or 2). And yet these undetected partial errors were corrected for, and were associated with an early ERN signal. Both data points suggest that the partial error was registered, and quickly corrected.

What candidate perceptual state are we saying is unconscious in these conditions? There are two options. First, the sensory registration of error is plausibly unconscious, given that participants reported no error, and were highly confident that no error has taken place. One might, cutting things rather fine, argue that sensory registration of error is a sensorimotor phenomenon, and not a matter of sensory perception. But error perception seems as much a genuine function of perception as is object perception, and is arguably a sub-type of event perception. Further, there is a second option here – the sensory representation of eye movement upon which the error registration depends. One might, of course, claim that the sensory representation of eye movement is conscious, while the error registration is not. But absent any reason to think so, this is special pleading. We submit that both states are good candidates for unconscious perception.

The second kind of study we wish to discuss involves typing. Consider the following paradigm, due to Logan and Crump (2010). Skilled⁷ typists would see a word on a screen, and then type the word. After typing it, they would see what purported to reflect what they typed beneath the target word. In reality Logan and Crump would occasionally insert correct responses when errors were made, and vice versa. Logan and Crump asked, in various ways, for the typists' judgments about whether they had made an error. And they tested for detection of error by looking for

⁷ Skill level was assessed on the basis of Word Per Minute (WPM) performance on a typing test. The participants were all college-aged and "typed at speeds comparable to professional typists" (p. 685).

evidence of post-error slowing. Note that in one way this is a strong test for error registration, since studies on partial error often find post-error slowing only in cases of awareness reports of the partial error.⁸

Let us focus on experiment three, which is the most important one for our purposes. Logan and Crump explicitly informed participants that in some cases errors would be corrected on the screen, and in others errors would be inserted instead of the correct response typists gave. The (skilled) typists typed 600 words each, with an average success rate of 91.8 percent. After each word they were given a choice between four options regarding what had happened on the screen: correct, error, inserted error, and corrected error.

[Table 1 here]

Correct Response	Error	Inserted Error	Corrected Error
No PES	PES	No PES	PES
Could discriminate from Corrected Errors	Could discriminate from Inserted Errors	Could discriminate from Errors	Could not discriminate from Correct Responses
Report matches typing behavior	Report matches typing behavior	Report matches typing behavior	Report does not match typing behavior

Table 1. An explanation of the results from study 3 of Logan and Crump (2010). The key condition is the rightmost, in which, although PES was generated, participant reports did not differ from the leftmost condition. In other words, reports tended to follow visual evidence of no error, even though typing behavior (PES) indicates registration of error.

⁸ There is a potential tension here, of course, in our use of saccade studies that find minimal PES for unaware errors and the typing study that finds PES for errors that, we will suggest, involve no conscious perception. In our view the tension is something that needs to be resolved by further research – for data on aware and unaware errors is not at present fine-grained enough to allow detailed conclusions regarding the role of awareness in generating the PES response. For all we know, unaware errors could generate PES in some circumstances.

Typists demonstrated reliable post-error slowing for actual errors, no matter what was shown on the screen. But they did not display post-error slowing following correct responses for which an error was inserted on the screen. So post-error slowing was reliably driven by error registration mediated by haptic representations of finger movement (perhaps, of course, in conjunction with (copies of) motor commands).

The question is whether these haptic representations of finger movement were conscious or unconscious. Logan and Crump found that while typists could discriminate (above chance) between actual errors and inserted errors—they said ‘error’ more than ‘inserted’ for actual errors and ‘inserted’ more than ‘error’ for inserted errors—they could not discriminate (above chance) between correct responses and corrected errors. When they had made an error, but the screen showed a correct response, they reported just as they did in cases of actual correct responses. And yet they would display PES in these cases, suggesting both sensory registration of error, as well as a lack of awareness of the registration of error.⁹

That is the bulk of the case for unconscious haptic perception in typing. The typists registered an action error, as measured by PES.¹⁰ But their reports suggested that they were unable to discriminate between correct responses and corrected errors. The report does not match the typing behavior. So – arguably – there was an unconscious sensory registration of error, as well as a haptic representation of finger movement upon which the error registration depends, that engendered the action guidance processes associated with error registration.

4.3 A response, and an interesting new direction

⁹ One might find the disparity between inserted error (success) and corrected error (error) cases puzzling. Note, however, that a similar pattern shows up in other studies of error awareness. People seem more accurate at reporting successes than errors.

¹⁰ Here measured by “examining interkeystroke interval for the trial before and two trials after an error” (p. 685).

An important feature of our case is that the phenomenon it draws on – sensory registration of error – is very broad, and can be studied in many different contexts. An important corollary is that unconscious sensory registration of error should be detectable via multiple experimental paradigms. We have discussed two. But the case leaning on the second (skilled typing) paradigm could, in our view, be stronger. This is in part because the experimenters running this line of research had different questions in mind. But studies on typing could be run that, for example, paid closer attention to confidence levels in reports of error awareness. And indeed, a possible response from Phillips points the way to an interesting line of research.

If we were Phillips, we would point out that – eyeballing the data Logan and Crump report – though typists were at chance in discriminating between actual correct responses and inserted correct responses (i.e., corrected errors), they were well above zero at noticing that a corrected error had taken place. Logan and Crump did not analyze this fact, so it is unclear whether the difference is significant. But, one might think, if the haptic registrations of error were fully and in every case unconscious, wouldn't the typists have sided with the screen in almost all cases? Wouldn't they have missed almost all of the corrected errors?

The possibility we wish to explore is that typist reports are informed by metacognitive feelings that result from performance monitoring (cf. Berger and Mylopoulos 2019). More research is needed to determine whether this is true, and what sort of metacognitive feelings may be present in the typing case. One option is that typists are here relying on a feeling of confidence in their performance. The lower the feeling of confidence the more likely they are to report that they have made an actual error. It is reasonable to think that their feelings of confidence in their performance would be stronger in the case of making an error that is then corrected via visual input vs. that in which they make an error that is not so corrected. In addition, it is interesting to note in this connection that Fairhurst et al. (2018) found that in conditions of multisensory conflict,

metacognitive confidence in touch representations were raised relative to confidence in visual representations.

A second option is that typist reports in inserted error cases are influenced by a feeling of fluency, perhaps resulting from the movements matching sensorimotor predictions (Fleming et al. 2015). Indeed, as Gajdos et al. (2019) note, recent work on metacognition suggests sensitivity to a wide range of evidence: ‘the entire perception-action cycle, and not only variation in sensory evidence, may be monitored when evaluating confidence in one’s decisions (Pouget *et al.* 2016; Fleming and Daw 2017). Such models lead to the counterintuitive prediction that confidence in perceptual judgements should be specifically modulated by features of motor output as well as perceptual input’ (Gajdos et al. 2019, 1). Gajdos et al. note a number of factors that could be relevant in the typing case under consideration. For example, response time modulates confidence in a perceptual judgment (Kiani et al. 2014). So does unexpected arousal (Allen et al. 2016). So does a change to the agent’s movement speed (Palser et al. 2018). Whether any of these factors are taken up by metacognition in the typing case is unknown, but the evidence is certainly suggestive. So agents could have a sense – informed by metacognition, and in particular by either an absence of or a lowered feeling of fluency in corrected error cases – that something is awry with how they typed, even if the sensory registrations of error driving PES are not conscious. And they may be more inclined to base their reports on these feelings in cases in which the screen matches what is indicated by them (error), versus cases in which it does not (inserted correction).

Now, if we are to appeal to metacognition to explain typist reports, rather than to perception, we need to think it plausible that metacognition utilizes unconscious states. We think there is some reason to do so.

Charles et al. (2013) had participants perform a task of comparing a number on a screen to the number 5. They had varying times to view the number – times of 16 ms, 33 ms, 50 ms, 66 ms,

and 100 ms – after which, in each case, the number was replaced by a mask. In each case they were asked to perform the comparison, and then to report both whether they had seen the number, and whether they thought they performed the comparison correctly.

Charles and colleagues compared trials on which participants reported they had seen the number, and trials on which they reported they had not seen the number. Unsurprisingly, there were important differences between the two sets of trials. Seen trials were associated with higher rates of correct judgments, as well as with a specific neural signal (the error-related negativity).

Importantly, however, participants were above chance at detecting the difference between correct and mistaken responses, even on unseen trials, which included viewing times of 16, 33, and 50 ms. Charles et al. comment: ‘participants performed above chance in detecting their own errors even on unseen trials. In both experiments, meta-cognitive performance on unseen trials increased with SOA, suggesting that longer SOAs allowed increasing amounts of evidence to be accumulated’ (2013, 90).¹¹

The upshot is that metacognition can, and does, operate over unconscious states (for further evidence of this, see Cortese et al. 2020). None of this proves, of course, that typist reports were reliant upon metacognition, as opposed to perception. But it is a tantalizing possibility that future experimentation could explore. Indeed, one thing that is genuinely exciting about this line of work is that it suggests a way forward on a debate that veers toward impasse. It is possible to study the kind of metacognitive feelings present in these cases, and to link data from such studies with other work on metacognition of unconscious processes. A cleaner account of how metacognition utilizes

¹¹ Look: the Charles et al. studies were not measuring whether access to these unconscious representations were useful for central coordinating agency. But it certainly seem to be. If agents have a sense of how well they are performing, they will be able to use this sense to tweak conditions of performance. In the good case, this will lead them to improve performance. So it is arguable that human agents have a kind of indirect access to unconscious representations via metacognition. Note that AVAILABILITY and EXPLOITABILITY, as we formulated them, made no mention of whether access should be direct or indirect. It seems to us one needs a special reason to favor directness here.

conscious and unconscious states and processes, and of the conditions – e.g., of multisensory conflict, and of multisensory integration (cf. Deroy et al. 2016) – that change metacognitive feelings and their accuracy, is in principle achievable. Taking account of the role of metacognition in such conditions seems an important part of the inquiry that the debate on unconscious perception helps motivate, namely, that of determining the role and status (whether perceptual, or conscious) of various sensory and sensorimotor states in the operation of central coordinating agency.

5 Conclusion

Thus far we have primarily done two things. First, we placed a more careful understanding of the attributability of a state to the individual in the context of results regarding perception and action guidance. This more careful understanding runs through the notion of central coordinating agency. We offered an explication of what it is for a state to be available to central coordinating agency.

Second, we offered novel interpretations of two studies. The first, concerning vision-for-action, has often been taken to support the view that unconscious visual states guide behavior on ‘auto-pilot.’ We argued that this interpretation is not mandatory, and that there is good reason to think that these states qualify as attributable to the individual. The second study concerned the role of ‘inner-loop’ processes in the guidance of typing. We argued that haptic representations that are a part of these processes qualify as attributable to the individual. Both interpretations place pressure on the unconscious perception skeptic. Further, the second type of study suggests that [a] an exploration of the role of haptic representation in action guidance is a promising avenue for further reflection, and [b] any full account of the role of unconscious and conscious states in the guidance of action requires greater attention to the role of metacognition.

In what remains we turn to two more broadly philosophical points.

First, the attention we have given to the nature of central coordinating agency interfaces in an interesting way with views regarding the nature – and more specifically the kindhood – of perception. Phillips contrasts a view on which perception is a natural psychological kind with one on which it is a manifest kind. If perception is a natural kind, as Burge (2010) and Block (in Phillips and Block 2016) and many others maintain, Phillips notes it is plausible to view consciousness as ‘part of the prototype used to identify instances of perception, leaving it as an open scientific question whether consciousness forms part of the essence of the kind’ (2018a, 475). That seems right, but it also seems right that if perception is a natural kind, one should have high credence in the existence of some unconscious perceptual states. This is because the natural kinds we find in biology and psychology are most plausibly homeostatic property clusters (Boyd 1991). Such clusters do not support lines as clean as the one required by the unconscious perception skeptic (cf. Taylor 2019). This skeptic suspects that every instance of perception is conscious. But even if consciousness is typical, or paradigmatic, of a perceptual state, the homeostatic property cluster view of natural kinds leads one to expect that some cases will exist which flout the trend. These might be cases just like the ones explored in this paper. This could be cases of vision operating to guide action at rapid timescales, or in fine-grained ways. Or this could be cases of haptic representations operating as a part of a representational mechanism built up and made efficient by years of training.

Phillips’s development of the idea that perception is rather a manifest kind is subtle. He does not deny that scientific psychology can uncover interesting facts about perception. But these would be facts about the psychological kind that constitutes the manifest kind. So the manifest kind can remain essentially conscious.

Call the kind identified by perceptual psychology, P. P should not be thought of as identical with, but as constituting, perception. Perception is rather a manifest kind. It is constituted by

P only when certain requirements of manifest form are met. Such requirements involve having a phenomenal nature. Hence all perception is conscious. This reply is quite consistent with the existence of a psychological science of perception whose kinds can occur unconsciously. (477)

One might worry about the support for the view that the requirements of perception's manifest form involve having a phenomenal nature. We wish to press a different point. The plausibility of the manifest kind view may turn in part on whether perception is *special*, as Phillips suggests (2018a, 476), in that it fails to have an underlying hidden nature not already worn 'on its sleeve' (475). Some questions: is perception alone special, or is cognition special? Is action special? Can we theorize about the nature of perception without theorizing about other psychological capacities?

We have illuminated the sense in which a view of perception depends upon ideas about agency. In particular, in order to understand what kinds of states qualify as perceptual, one needs to understand the nature of central coordinating agency. And so it seems that the proponent of a view that perception is a manifest kind is taking on commitments that go beyond the nature of perception. We have articulated a perspective on central coordinating agency. We do not know whether central coordinating agency is a natural kind, but it is not plausible that it is a manifest kind. Phillips does not go so far as to claim that perception essentially involves links to central coordinating agency – the link he posits is evidential. But even so, this link undermines the claim that 'perception lacks a hidden, underlying nature' (2018a, 476). For in some cases the best way to tell whether a state is perceptual is to examine its links to something that does have such an underlying nature, namely, central coordinating agency.

A second issue is closely related. We have offered an explication of attributability to the individual, and it has done work in our argumentation. But we are open to the possibility that there are other fruitful explications of attributability, or indeed, of availability to central coordinating agency (see, e.g., Buehler 2019). This raises the possibility that on some explications of attributability, or of availability to central coordinating agency, an unconscious state qualifies, while on other explications, the same state fails to qualify. And that pushes one to accept the possibility that the same state-type could be classified as perceptual by one theorist and as sub-perceptual by another theorist, even though both theorists are working with fruitful and plausible understandings of what it is that makes a state genuinely perceptual.

Phillips is aware of the situation. We have already seen that he sets aside Evans's understanding of attributability to avoid begging the question regarding unconscious perception. Now consider two interesting points Phillips acknowledges in a footnote (2018a, fn. 45). First, Burge (2010) indicates that he rejects what we have called Idea Two (335, fn.62), rendering a Burgean explication of attributability not dependent upon guidance, and thus in many cases weaker than Phillips's. Second, Bayne (2013) offers an explication of attributability based upon information integration that seems even stronger than Phillips's. Bayne's is 'one which would rule out many representations – including even those implicated in familiar cases of blindsight – from counting as individually attributable' (Phillips 2018a, 499).

What should we say? Phillips is invoking a notion that is supposed to do the work of blocking inferences to the existence of unconscious perception. But it seems there is significant disagreement as to what this notion amounts to. One may be tempted to agree with Block.

While there are clear cases of sub-personal representations (such as gastrointestinal representations) and personal representations (e.g., conscious perceptions), many if not most

cases of interest are indeterminate and there is no accepted characterization of the difference. Every proposal that has been made for what the personal/sub-personal distinction comes to has an air of postulation. (Peters et al. 2017, 8)

Is this whole dispute *merely verbal*?

We would agree with Phillips that it is not. Certainly the dispute has several moving parts. And it seems like a live possibility that there exist unconscious states the perceptual status of which depends upon some theorist's explication. But in our view discovering the dependence relations would qualify as genuine philosophical progress. Once we see how an unconscious state-type is related to the features that factor in some explication of attributability, we will be in a better position to fit that state into broader accounts of perception or of sub-perceptual processing. And we will be in a better position to determine the suitability of this state to play roles in ancillary issues that give this debate so much punch – issues we mentioned in this paper's first paragraph, like the epistemic role of perception, the functions of perception, and the functions of consciousness.

Of course parts of this dispute could turn on disagreements that are merely verbal. We agree with Chalmers (2011) that discovery of verbal disputes – and discovery of the sources of verbal disputes – can in itself constitute philosophical progress. A benefit of clear explications of attributability, like the one we have developed here, is that we can understand what is at stake. For that reason we encourage those who find fault with our explication of attributability to offer their own. The governing attitude is Shakespearean: Let us not to the marriage of true minds admit impediments.

Funding

Joshua Shepherd acknowledges (with thanks) funds from European Research Council Starting Grant 757698, awarded under the Horizon 2020 Programme for Research and Innovation, as well as a fellowship from the Canadian Institute for Advanced Research's Azrieli Global Scholar programme, and the Brain, Mind, and Consciousness program. Myrto Mylopoulos is supported in part by funding from the Social Sciences and Humanities Research Council of Canada (Grant No. 430-2017-00811).

Acknowledgements

We wish to thank a perceptive referee, and Ian Phillips for helpful comments on an earlier draft.

References

- Allen, Micah, Darya Frank, D. Samuel Schwarzkopf, Francesca Fardo, Joel S. Winston, Tobias U. Hauser, and Geraint Rees (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *Elife*, 5, e18103.
- Amengual, J. L., J. Marco-Pallares, L. Richter, S. Oung, A. Schweikard, B. Mohammadi, A. Rodriguez-Fornells, and T. F. Münte (2013). Tracking post-error adaptation in the motor system by transcranial magnetic stimulation. *Neuroscience* 250: 342-351.
- Bayne, Timothy (2013) Agency as a marker of consciousness. *Decomposing the Will*. eds. T. Vierkant, J. Kiverstein, and A. Clark. Oxford University Press, 160–182.
- Barceló, Francisco, José A. Periañez, and Robert T. Knight (2002). Think differently: A brain orienting response to task novelty. *NeuroReport* 13(15): 1887-1892.

Berger, Jacob, & Mylopoulos, Myrto (2019). On scepticism about unconscious perception. *Journal of Consciousness Studies*, 26(11-12), 8-32.

Berger, Jacob, & Mylopoulos, Myrto (Forthcoming). Default hypotheses in the study of perception: A reply to Phillips. *Journal of Consciousness Studies*.

Berger, Jacob, & Nanay, Bence (2016). Relationalism and unconscious perception. *Analysis*, 76(4), 426-433.

Berger, Jacob, Nanay, Bence, & Quilty-Dunn, Jake (2018). Unconscious perceptual justification. *Inquiry*, 61(5-6), 569-589.

Botvinick, Matthew M., Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen (2001). Conflict monitoring and cognitive control. *Psychological review*, 108(3), 624.

Boyd, Richard (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 127-148.

Bridgeman, Bruce, Stephen Lewis, Gary Heit, and Martha Nagle (1979). Relation between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology: Human Perception and Performance*, 5(4), 692.

Brozzo, Chiara (2017). Motor intentions: How intentions and motor representations come together. *Mind & Language*, 32(2), 231-256.

Buehler, Denis. (2019). Flexible occurrent control. *Philosophical Studies*, 176, 2119–2137.

Burge, Tyler (2010). *Origins of objectivity*. Oxford University Press.

Burnston, Daniel C. (2017). Interface problems in the explanation of action. *Philosophical Explorations*, 20(2), 242-258.

Butterfill, Stephen A., & Sinigaglia, Corrado (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, 88(1), 119-145.

Cameron, Brendan D., Erin K. Cressman, Ian M. Franks, and Romeo Chua (2009). Cognitive constraint on the 'automatic pilot' for the hand: movement intention influences the hand's susceptibility to involuntary online corrections. *Consciousness and cognition*, 18(3), 646-652.

Campbell, John (2002). *Reference and consciousness*. Oxford University Press.

Campbell, John (2004). The first person, embodiment, and the certainty that one exists. *The Monist* 87(4): 475-488.

Castiello, Umberto, Yves Paulignan, and Marc Jeannerod (1991). Temporal dissociation of motor responses and subjective awareness: A study in normal subjects. *Brain*, 114(6), 2639-2655.

Ceccarini, Francesco, and Umberto Castiello (2018). The grasping side of post-error slowing. *Cognition*, 179, 1-13.

Chalmers, David J. (2011). Verbal disputes. *Philosophical Review*, 120(4), 515-566.

Charles, Lucie, Filip Van Opstal, Sébastien Marti, and Stanislas Dehaene (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, 73, 80-94.

Chiu, Yu-Chin, and Tobias Egner (2015). Inhibition-induced forgetting results from resource competition between response inhibition and memory encoding processes. *Journal of Neuroscience* 35(34): 11936-11945.

Christensen, Wayne, Bicknell, Kath, McIlwain, Doris, & Sutton, Jon (2015). The sense of agency and its role in strategic control for expert mountain bikers. *Psychology of Consciousness: Theory, Research, and Practice*, 2(3), 340.

- Christensen, Wayne, Sutton, Jon, & McIlwain, Doris J. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language*, 31(1), 37-66.
- Clark, A. (2001). Visual experience and motor action: Are the bonds too tight? *Philosophical Review*, 110, 495–519.
- Cortese, Aurelio, Hakwan Lau, and Mitsuo Kawato (2019). Metacognition facilitates the exploitation of unconscious brain states. *BioRxiv*, 548941.
- Danielmeier, Claudia, Tom Eichele, Birte U. Forstmann, Marc Tittgemeyer, and Markus Ullsperger (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *Journal of Neuroscience* 31(5): 1780-1789.
- Day, Kevin A., Ryan T. Roemmich, Jordan A. Taylor, and Amy J. Bastian (2016). Visuomotor learning generalizes around the intended movement. *eNeuro*, 3(2).
- Dennett, Daniel C. (2002). *Content and consciousness*. Routledge.
- Deroy, Ophelia, Spence, Charles, & Noppeney, Uta (2016). Metacognition in multisensory perception. *Trends in cognitive sciences*, 20(10), 736-747.
- Drayson, Zoe (2014). The personal/subpersonal distinction. *Philosophy Compass*, 9(5), 338-346.
- Dutilh, Gilles, Joachim Vandekerckhove, Birte U. Forstmann, Emmanuel Keuleers, Marc Brysbaert, and Eric-Jan Wagenmakers (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, 74(2), 454-465.
- Endrass, Tanja, Benedikt Reuter, and Norbert Kathmann (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience* 26, 6: 1714-1720.
- Evans, Gareth (1982). *The varieties of reference*, J McDowell (ed.). Oxford: Clarendon Press.

- Fairhurst, Merle T., Eoin Travers, Vincent Hayward, and Ophelia Deroy (2018). Confidence is higher in touch than in vision in cases of perceptual ambiguity. *Scientific reports*, 8(1), 1-9.
- Ferretti, Gabriele, & Caiani, Silvani Z. (2019). Solving the interface problem without translation: The same format thesis. *Pacific Philosophical Quarterly*, 100(1), 301-333.
- Ficarella, Stefania C., Nicolas Rochet, and Boris Burle (2019). Becoming aware of subliminal responses: an EEG/EMG study on partial error detection and correction in humans. *Cortex* 120: 443-456.
- Fleming, Stephen M., & Daw, Nathaniel D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological review*, 124(1), 91.
- Fleming, Stephen M., Brian Maniscalco, Yoshiaki Ko, Namema Amendi, Tony Ro, and Hakwan Lau (2015). Action-specific disruption of perceptual confidence. *Psychological science*, 26(1), 89-98.
- Frankfurt, Harry G. (1978). The problem of action. *American philosophical quarterly*, 15(2), 157-162.
- Fridland, Ellen (2017). Automatically minded. *Synthese*, 194, 4337-4363.
- Fridland, Ellen (2017). Skill and motor control: intelligence all the way down. *Philosophical Studies*, 174(6), 1539-1560.
- Fridland, Ellen (2019). Intention at the Interface. *Review of Philosophy and Psychology*, 1-25.
- Fleming, Stephen, Marta Saez Garcia, Gabriel Weindel, and Karen Davranche (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of consciousness*, 2019(1), niz001.
- Gajdos, Thibault, Stephen M. Fleming, Marta Saez Garcia, Gabriel Weindel, and Karen Davranche (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of consciousness* 2019(1): niz001.

- Goodale, Mel A., & Humphrey, G. Keith (1998). The objects of action and perception. *Cognition*, 67(1-2), 181-207.
- Goodale, Melvyn A., Denis Pelisson, and Claude Prablanc (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320(6064), 748-750.
- Jiang, Yi, Patricia Costello, Fang Fang, Miner Huang, and Sheng He (2006). A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences*, 103(45), 17048-17052.
- Kiani, Roozbeh, Leah Corthell, and Michael N. Shadlen (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, 84(6), 1329-1342.
- Kopiske, Karl K., Nicola Bruno, Constanze Hesse, Thomas Schenk, and Volker H. Franz. "The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study." *Cortex* 79 (2016): 130-152.
- Krakauer, John W., Alkis M. Hadjiosif, Jing Xu, Aaron L. Wong, and Adrian M. Haith (2019). Motor learning. *Comprehensive Physiology*, 9(2), 613-663.
- Liu, Xianyun, Matthew J. Crump, and Gordon D. Logan (2010). Do you know where your fingers have been? Explicit knowledge of the spatial layout of the keyboard in skilled typists. *Memory & cognition*, 38(4), 474-484.
- Logan, Gordon D., & Crump, Matthew J. (2009). The left hand doesn't know what the right hand is doing: The disruptive effects of attention to the hands in skilled typewriting. *Psychological Science*, 20(10), 1296-1300.

- Logan, Gordon D., & Crump, Matthew J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, 330(6004), 683-686.
- Milner, David, & Goodale, Melvyn (2006). *The visual brain in action*. Oxford University Press.
- McDougle, Samuel D., Ivry, Richard B., & Taylor, Jordan A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in cognitive sciences*, 20(7), 535-544.
- Mole, Christopher (2009). Illusions, demonstratives and the zombie action hypothesis. *Mind*, 118(472): 995-1011.
- Moors, Agnes (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, 67, 263-287.
- Moors, Agnes., & De Houwer, Jan (2006). Automaticity: a theoretical and conceptual analysis. *Psychological bulletin*, 132(2), 297-326.
- Mylopoulos, Myrto, & Pacherie, Elisabeth (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology*, 8(2), 317-336.
- Mylopoulos, Myrto, & Pacherie, Elisabeth (2019). Intentions: The dynamic hierarchical model revisited. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(2), e1481.
- Norman, Liam J., Charles A. Heywood, and Robert W. Kentridge (2013). Object-based attention without awareness. *Psychological science*, 24(6), 836-843.
- Notebaert, Wim, Femke Houtman, Filip Van Opstal, Wim Gevers, Wim Fias, and Tom Verguts (2009). Post-error slowing: an orienting account. *Cognition*, 111(2), 275-279.
- Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts metacognitive accuracy. *Consciousness and cognition*, 57, 33-40.

- Peters, Megan A., & Lau, Hakwan (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *Elife*, 4, e09651.
- Peters, Megan A., Ro, Tony, & Lau, Hakwan (2016). Who's afraid of response bias?. *Neuroscience of consciousness*, 2016(1), niw001.
- Peters, Megan A., Kentridge, Robert W., Phillips, Ian, & Block, Ned (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of consciousness*, 3(1).
- Phillips, Ian (2016). Consciousness and criterion: on Block's case for unconscious seeing. *Philosophy and Phenomenological Research*, 93(2), 419-451.
- Phillips, Ian (2018). Unconscious perception reconsidered. *Analytic Philosophy*, 59(4), 471-514.
- Phillips, Ian (2018). The methodological puzzle of phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755), 20170347.
- Phillips, Ian (2019). Object files and unconscious perception: a reply to Quilty-Dunn. *Analysis*.
- Phillips, Ian (Forthcoming). Scepticism about unconscious perception is the default hypothesis. *Journal of Consciousness Studies*.
- Phillips, Ian, & Block, N. (2016). Debate on unconscious perception. *Current controversies in philosophy of perception*, 165-192.
- Pisella, Laure, H. Grea, C. Tilikete, A. Vighetto, Michel Desmurget, Gea Rode, D. Boisson, and Yves Rossetti (2000). An 'automatic pilot' for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. *Nature neuroscience*, 3(7), 729-736.
- Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature neuroscience*, 19(3), 366.

- Posner, Michael, I. & Snyder, Charles, R. (1975). Attention and cognitive control. In R. L. Solso (ed.) *Information processing and cognition: the Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.
- Quilty-Dunn, Jake (2019). Unconscious perception and phenomenal coherence. *Analysis*, 79(3), 461-469.
- Regev, Shirley, and Nachshon Meiran (2014). Post-error slowing is influenced by cognitive control demand. *Acta psychologica*, 152, 10-18.
- Rochet, Nicolas, Laure Spieser, Laurence Casini, Thierry Hasbroucq, and Boris Burle (2014). Detecting and correcting partial errors: Evidence for efficient control without conscious access. *Cognitive, Affective, & Behavioral Neuroscience* 14, no. 3: 970-982.
- Shepherd, Joshua (2015). Conscious control over action. *Mind & language*, 30(3), 320-344.
- Shepherd, Joshua (2016) Conscious action/Zombie action. *Noûs*, 50(2), 419–444.
- Shepherd, Joshua (2019). Skilled action and the double life of intention. *Philosophy and phenomenological research*, 98(2), 286-305.
- Shepherd, Joshua (2021). Skill and sensitivity to reasons. *Review of Philosophy and Psychology*, 1-13.
- Shiffrin, Richard M., & Schneider, Walter (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review*, 84(2), 127.
- Siegel, Susanna (2016). *The rationality of perception*. Oxford University Press.
- Smithies, Declan (2019). *The epistemic role of consciousness*. Oxford University Press.
- Snodgrass, Michael, & Shevrin, Howard (2006). Unconscious inhibition and facilitation at the objective detection threshold: Replicable and qualitatively different unconscious perceptual effects. *Cognition*, 101(1), 43-79.

- Taylor, Henry (2019). Fuzziness in the mind: Can perception be unconscious? *Philosophy and Phenomenological Research*.
- Taylor, A. J. G., and S. B. Hutton (2011). Error awareness and antisaccade performance. *Experimental brain research* 213, no. 1: 27-34.
- Tian, Xiaoguang, Masatoshi Yoshida, and Ziad M. Hafed (2018). Dynamics of fixational eye position and microsaccades during spatial cueing: the case of express microsaccades. *Journal of neurophysiology* 119, 5: 1962-1980.
- Ullsperger, Markus, Helga A. Harsay, Jan R. Wessel, and K. Richard Ridderinkhof (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function* 214(5): 629-643.
- Ullsperger, Markus, & Danielmeier, Claudia (2016). Reducing speed and sight: how adaptive is post-error slowing? *Neuron*, 89(3), 430-432.
- Wessel, Jan R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, 55(3), e13041.
- Wessel, Jan R., and Aron, Adam R. (2017). On the globality of motor suppression: unexpected events and their influence on behavior and cognition. *Neuron* 93(2): 259-280.
- Willeke, Konstantin F., Xiaoguang Tian, Antimo Buonocore, Joachim Bellet, Araceli Ramirez-Cardenas, and Ziad M. Hafed (2019). Memory-guided microsaccades. *Nature communications* 10 (1) (2019): 1-14.
- Wolpert, Daniel M. (1997). Computational approaches to motor control. *Trends in cognitive sciences* 1(6): 209-216.
- Wu, Wayne (2011). Confronting Many-Many problems: Attention and agentive control. *Notes*, 45(1), 50-76.

Wu, Wayne (2013a). The case for zombie action. *Mind*, 122(485), 217–230.

Wu, Wayne (2013b). Mental action and the threat of automaticity. In Andy Clark, Julian Kiverstein & Tillman Vierkant (eds.), *Decomposing the Will*. Oxford University Press. 244-61.

Wu, Wayne (2014). *Attention*. Routledge.

Wu, Wayne (2016). Experts and deviants: The story of agentic control. *Philosophy and Phenomenological Research*, 92(2), 101-126.

Wu, Wayne (2020). Is vision for action unconscious? *Journal of Philosophy*, 117(8), 413-433.