



# Perception needs modular stimulus-control

Anders Nes<sup>1</sup>

Received: 4 October 2022 / Accepted: 30 April 2023 / Published online: 25 May 2023  
© The Author(s) 2023

## Abstract

Perceptual processes differ from cognitive, this paper argues, in functioning to be causally controlled by proximal stimuli, and being modular, at least in a modest sense that excludes their being isotropic in Jerry Fodor's sense. This claim agrees with such theorists as Jacob Beck and Ben Phillips that a function of stimulus-control is needed for perceptual status. In support of this necessity claim, I argue, inter alia, that E.J. Green's recent architectural account misclassifies processes deploying knowledge of grammar as perceptual. *Pace* Beck and Phillips, however, I argue a function of stimulus-control is insufficient for perceptual as opposed to cognitive status. One consideration in favour of such insufficiency, noted but (I argue) not convincingly rebutted by these theorists, concerns perpetually grounded demonstrative thought. Two other considerations trade on the fact that a function of stimulus-control can arise not from blind nature but intentional design or social institutions, where so-functioning processes may but need not be perceptual. I offer two cases where such processes are cognitive, viz. skilful play-by-play announcing of ongoing events, and voluntary visualizing of ongoing events under the guidance of apt play-by-play announcements, dubbed announcement-driven visualizing (ADV). The cognitive status of these three diverse phenomena cannot be explained by an absence of a perception-like representational format or content (for ADV has such) or by a presence of personal-level mental states causally mediating between stimuli and outputs (for perception has such). A better explanation invokes, I argue, the non-modular character of the generating process.

**Keywords** The perception/cognition distinction · Imagery · Modularity

---

✉ Anders Nes  
anders.nes@ntnu.no

<sup>1</sup> Department of Philosophy and Religious Studies, NTNU–Norwegian University of Science and Technology, Dragvoll Alle 40, 7491 Trondheim, Norway

## 1 Introduction

The idea that there is a difference between perception and thought, or perception and cognition as it is often put, is entrenched in philosophy, psychology, and common sense.<sup>1</sup> Yet the question just *how* perception and cognition differ has long been curiously neglected, despite its obvious interest for understanding the ground of their different epistemic roles. Philosophical attention to the issue is growing however with theorists recently pointing to phenomenological differences (e.g., Kriegel, 2019), dissimilarities in representational format or content (e.g., Block, 2023), or junctures in cognitive architecture (e.g., Firestone & Scholl, 2016; Green, 2020) as important contrasts (for review, see Nes et al., 2023).

Another such putative difference is the dependence of perception on (proximal) stimuli: light hitting the retina, pressure waves oscillating the cochlea, and so on. The idea that perception characteristically is stimulus-dependent fits with the standard view in perceptual psychology that perceptual processes function to take stimuli as input and ‘pick up’, extract, or compute therefrom information about the surroundings (cf. Gibson, 1966, pp. 1–5; Marr, 1982, pp. 3–7; Rock, 1983, pp. 29, 98; Palmer, 1999, pp. 5–6; Burge, 2010, p. 89). More broadly, since the senses function to respond to a corresponding type of stimulus, it chimes with the common-sense notion that perception has to do with the use of the senses. Recently, Jacob Beck (2018) and Ben Phillips (2019, 2021), have defended this approach to the perception/cognition-distinction in detail, arguing that perceptual states differ from cognitive states in having the function of being causally sustained or controlled by stimuli.<sup>2</sup>

This paper agrees that a function of stimulus-dependence, or stimulus-control (to use the term I shall prefer, for reasons noted in Sect. 2), is needed for a process to be perceptual as opposed to cognitive. In support of this claim, I argue, among other things, that Green’s (2020) recent ‘dimension restriction hypothesis’ fails to account for why cognitive processes deploying knowledge of grammar are not perceptual. However, *pace* Beck and Phillips, I doubt a function of stimulus-control is enough for perceptual as opposed to cognitive status. Three interestingly different cases of cognitive processes functioning to be stimulus-controlled will be offered.

The first case, acknowledged but (I argue) inadequately handled by Beck and Phillips, is perceptual-demonstrative thought. The two other problems result from the fact that functions can arise in sundry ways, not just from nature but via intentional design or social institutions and in the skills with which they may correspond. We cannot simply set aside functions arising non-naturally in characterising perception, for such functions can arguably be vital to perception, say for perceptual systems in robots. However, although human inventiveness may allow us to design processes with a bona fide perceptual status in which stimuli function to causally control distal representations, it also allows us to institutionalise or design processes with a putatively

---

<sup>1</sup> ‘Cognition’ in this context means extra-perceptual cognition, paradigmatically including thinking, reasoning, problem-solving and planning. Contrast various broader uses of ‘cognition’, e.g. in ‘cognitive science’.

<sup>2</sup> The approach is also suggested by Prinz (2006), Camp (2009), Burge (2010, p. 378), and Nanay (2015, p. 1729). Beck (2018, p. 320, nt. 1) traces it back to psychophysicists Fechner and Weber, and to Locke (1689/1979: II.xxxi.2).

*cognitive* status in which stimuli function to causally control distal representations. An example is the skilled activity of play-by-play announcing of unfolding events, e.g. football matches. Here, verbal representations, delivered by the announcer, function to be causally controlled by what hits her eyes.

The third case of a cognitive process functioning to be stimulus-controlled is more vision-like. It is a process designed to allow one to follow unseen events as they happen in visually immersed way, viz. by having visual imagery guided and causally controlled by apt play-by-play announcements. I dub this ‘announcement-driven visualizing’ (ADV). ADV adds interestingly to the two preceding problems in that it shows that it still will not do, to distinguish perception from cognition, to take the former to combine a function of stimulus control with a perception-like representational format or content.

Why are not perceptual-demonstrative thinking, play-by-play announcing, or ADV perceptual? Is it because, in each of these cases, stimuli cause the respective outputs by first causing certain intervening personal-level mental states that in turn cause the relevant outputs? However, as powerfully argued by such perceptual psychologist as Irvin Rock (1983, pp. 283–299, 1997, pp. 5–15), even paradigmatically perceptual processes, outputting e.g. perceptual representations of movement, may involve such intervening personal-level mental states. A better explanation of why our three cases are not perceptual, I argue, invokes a certain lack of modularity. Specifically, in each case, the outputs are generated by processes having an isotropic character, in Fodor’s (1983) sense, in that the processes have access, in their normal operation, to an open-ended variety among the attitudes that constitute the agent’s overall outlook. Their outputs are therefore fit to be regarded as cases of ‘what the agent makes of what her senses put to her’, intuitively speaking. Perceptual processes are modular, at least in a weak sense of being non-isotropic. This involves a comparative isolation from the agent’s outlook. Perceptions thereby remain, again in a manner of speaking, the work of the senses.

The next section introduces the stimulus-control approach, as recently defended by Beck and Phillips. Section 3 motivates the approach; specifically, the claim that a function of stimulus-control is needed for perceptual status. Section 4 rejects Beck’s and Phillips’s treatment of the problem of perceptual-demonstrative thought. Section 5 presents the problem of play-by-play announcing, while Sect. 6 answers some rejoinders. Sections 7 and 8 do the same for ADV, the latter among other things contrasting ADV with use of sensory substitution devices (SSDs). Section 9 argues that non-modularity—specifically, isotropy—best explains the non-perceptual status of our three cognitive processes functioning to be stimulus-controlled. Section 10 wraps up, observing an affinity with the views of Rock.

## 2 The stimulus-control approach

The category of the perceptual on which Beck and Phillips focus is one that includes some hallucinations (e.g. some involving spontaneous activations of visual cortices), yet excludes visualization and mental imagery, treating the latter as cognitive. The supposition that there is an interesting category of the perceptual along these lines

is standard in philosophy.<sup>3</sup> I will take it on board here. Beck and Phillips argue the uncoupling from stimuli in perceptual hallucinations amounts to a malfunction in the perceptual process.

Setting aside a subtlety to which we return in Sect. 4 below, Beck puts his thesis so:

**S-D FUNCTION:**  $\Psi$  is perceptual if, necessarily, all occurrences of  $\Psi$  have the function of being stimulus-dependent; otherwise,  $\Psi$  is cognitive. (Beck, 2018: 326, boldface added)

Here, ' $\Psi$ ' ranges over perceptual or cognitive state or event types. An occurrence of such a type is 'stimulus-dependent', in the relevant sense, 'just in case it is causally sustained by present proximal stimulation.' (Beck, 2018, p. 323).

A central thesis of Phillips's, dubbed 'Stimulus-Control<sub>p/C</sub>' is this:

**[Stimulus-Control<sub>p/C</sub>]** [A] process is perceptual just in case it has the function of producing representations of environment entities by being causally controlled by those proximal stimuli that these entities produce. (Phillips, 2019, p. 322)

The notion of causal control here is adopted from Stegmann (2014), who uses it to characterise how the sequence of amino acids assembled into proteins causally reflects the sequence of base pairs in DNA. Another paradigm example, also from Stegmann (2014:453), is how the sequence of tones played by a music box reflects the sequence of studs on the roll. To say perceptions are causally controlled by stimuli is to say their sequence similarly reflects a sequence of stimuli. More formally, a perceptual process exhibits causal control by stimuli, or, for short, stimulus-control, iff to the sequence of representational states generated by the process, ( $P_1, P_2, \dots$ ), there correspond to a sequence of stimuli, ( $S_1, S_2, \dots$ ), such that  $S_1$  is a cause of  $P_1$ ,  $S_2$  a cause of  $P_2$ , etc.

The notion of causal control has the advantage over that of causal sustainment of evidently permitting delay between stimuli and perceptual state. After all, can easily take a tenth of a second or more for signals from impacts upon pressure receptors in a toe even to reach the brain (cf., e.g. Siegel & Sapru, 2006, p. 257), and some milliseconds of cortical processing is needed for anything recognisable as a perceptual state to obtain. Impacts on the toe do not, then, sustain tactile perceptions as beams sustain a roof. Beck's notion of sustainment could probably be unpacked consistently with this point (cf. how a sequence of puffs sustains a feather in the air), but I will prefer the notion of causal control; hence the moniker 'stimulus-control approach'.

Whereas Phillips's Stimulus-Control<sub>p/C</sub> classifies processes producing representations Beck's S-D FUNCTION classifies representational states or events. As Beck (2018, p. 327) notes, though, it is plausible to think that if occurrences of  $\Psi$  have the function of being stimulus-dependent, this is explained by the mechanisms or processes engendering  $\Psi$ s: it is because these processes function to produce stimulus-dependent representations. The focus here will accordingly be on the processes. I will

<sup>3</sup> Cf. such representative overviews of philosophical work on perception as Fish (2010) and Pautz (2021), in both of which hallucination is discussed in detail but imagery at most in passing.

assume psychological states are perceptual rather than cognitive iff they are produced by perceptual as opposed to cognitive processes.

While both Beck's and Phillips's leading examples of functions are natural functions, such as the heart's function of pumping blood, they impose no limit to functions thus bestowed by the blind forces of nature (cf. Beck, 2018, p. 327, Phillips, 2019, pp. 320–321). This is well-advised, since it is plausible to think that artificial perceptual systems, in robots, or perhaps implanted into humans or animals, can in principle be produced. Whether such systems operate properly may well be dependent on the intentions of the designers or users.

Besides Stimulus-Control<sub>P/C</sub>, Phillips also defends 'Stimuli-Specific-Control<sub>P/C</sub>', a thesis designed to distinguish a narrower, sense-modality-specific notion of the perceptual. It differs from Stimulus-Control<sub>P/C</sub> in imposing the requirement upon perceptual processes (in this narrower sense) that they have the function of being controlled by stimuli specific to sense modality (or a specific mix of modality-specific stimuli). The primary focus here will be on Stimulus-Control<sub>P/C</sub>. I shall however argue (in Sect. 5) that not even a function of sense-modality-specific stimulus-control is enough for perceptual as opposed to cognitive status.

### 3 Stimulus-control needed

This section motivates the stimulus-control approach; specifically, the claim that a function of stimulus-control is needed for perceptual as opposed to cognitive status. I draw on points made by Beck and Phillips, whilst also updating or supplementing their case; in particular, I present a reason for dissatisfaction with Green's recent architectural proposal.

We noted two attractions of the stimulus-control approach: it fits with the common-sense notion that perception has to do with the use of the senses and coheres with the pervasive view in perceptual psychology that perceptual processes function to take stimuli as input and extract therefrom information about the surroundings. Moreover, the leading alternative grounds for differentiating perception from cognition do not seem adequate on their own, without adverting to stimulus-control.

Consider, first, invoking a non-discursive, iconic, or analogue representational format, or a non-conceptual or non-propositional content, to distinguish perception from cognition. One worry here is that there are reasons for thinking perception is at least in part discursive in format or conceptual/propositional in content (cf. Rock, 1983, pp. 43–99, Mandelbaum, 2018, Quilty-Dunn, 2020). A second concern is that visual imagery may seem to have the same type of format or content as vision (and likewise imagery corresponding to other modalities). Besides phenomenological similarities, perceptual and imagery states have overlapping functional profiles. They may, e.g., affect perceptual processing in similar ways. Thus, just as hearing a sound of collision can disambiguate an ambiguous 'stream or bounce' display into a 'bump' percept, so can auditory imagery of collision (Berger & Ehrsson, 2013). Moreover, imagery elicits eye movements akin to vision (cf. Laeng et al., 2014) and has overlapping neural bases (Kosslyn, 2005). Thirdly, nonconceptual contents, or iconic/analogue

formats, arguably have a role in rudimentary action-guiding, planning, or memory states (Burge, 2010, 2014).

A second option is to distinguish perception from cognition by a distinctive perceptual phenomenology. Firstly, however, this will not separate unconscious perception and cognition. Secondly, even some friends of a phenomenological criterion have found it is hard to say much informative about what the phenomenological difference between perception and cognition in general is (cf. Kriegel, 2019). Thirdly, and relatedly, at least some visual phenomenology may be pretty much like that of vivid visualization (cf. Phillips, 2019). Fourthly, if, contrary to the first concern (and to conventional wisdom) there is no unconscious perception, perception having faint, degraded phenomenology even in, say, blindsight subjects who evidence perceiving yet claim not to see anything (cf. Phillips, 2016), the third problem will be exacerbated, since perceptual phenomenology would subsume faint, unlively forms hard to distinguish from imagery phenomenology.<sup>4</sup>

A classic architectural criterion for distinguishing perception from cognition is cognitive impenetrability (cf. Firestone & Scholl, 2016). Now, though the debate on cognitive impenetrability is far from settled, various influential challenges remain, e.g. concerning whether attention- or expectation-mediated cognitive effects on perception amount to cognitive penetration (cf. Green, 2020; Mole, 2015; Stokes, 2018).

Recently, an alternative architectural difference between perception and cognition has been proposed by Green (2020). He offers a ‘dimension restriction hypothesis’, according to which perceptual but not cognitive processes are ‘dimensionally restricted’, in the sense that cognition cannot add to the dimensions that perceptual processes can represent and take into account in their computations. For example, cognitively appreciating this difference between being indoors or outdoors cannot make perceptual processes compute over the indoors/outdoors-dimension. Perception can take this dimension into account only if a sensitivity to the dimension in perceptual processes is either present at birth, emerges in maturation, or can be acquired through laborious processes of perceptual learning. However, even if cognition cannot add to the dimensions perceptual processes take into account, it can still affect what *value* they output on a given dimension, such as orientation or hue. Therefore, dimensional restriction permits cognitive penetration.

Now, at first blush, it is unclear why this information-processing profile of dimensional restrictedness should have anything distinctive to do with the use of the senses. Even if it were to turn out to be true, as a matter of lawful psychological fact, that only perceptual processes have the profile, one might have a lingering sense that the link with the *sensory* nature of perception is left obscure. More importantly, some cognitive processes arguably share the relevant information-processing profile.

---

<sup>4</sup> Interestingly, some influential attempts to spell out the distinctive ‘presentational’ phenomenology of perception have discerned a certain feeling in perception as of the objects of awareness causing one’s very unfolding awareness of these objects (Searle, 1983; Recanati, 2007, p. 133). The very phenomenology implies, on these views, that things are as it should with perception only if caused by the objects before us. This causation would (as a matter of basic psychological fact, not of phenomenology) be by way of stimuli. On these views, then, phenomenology, cum basic psychological fact, supports the centrality to perception of a function of stimulus-dependence.

Consider processes of deploying knowledge of grammar, or of updating that knowledge in first language acquisition. These processes are arguably architecturally restricted to computing over a fixed set of grammatical categories, to which general cognition cannot add. For example, although people cognitively appreciate, and care about, ordinal numbers (being first, or second, or third, etc.), in no known language is grammaticality contingent on the ordinal number position of words in the sentence, and it is doubtful whether such a language would even be psychologically possible (cf. Smith et al., 1993). Likewise, though people know and care about such dimensions as colour and symmetry, in no known language is grammaticality sensitive to such dimensions (Talmy, 1985, p. 134, Cinque, 2013). David Adger suggests, more generally:

Cultural concepts that may strongly influence how speakers of a language live their lives every day are not co-opted into grammar. The way that a culture dissects the world can be embedded in the words of a language, in the myths told in the language, in idioms and turns of phrase, but grammar, and phonology, are disconnected from worldview. (Adger, 2018, p. 29)

Processes deploying or updating knowledge of grammar are, on the face of it, cognitive. Chomsky, the main source of theorizing of grammar in the generative school, introduced the technical term ‘cognize’ to describe speakers’ grip on grammar. In doing so, he clarified that ‘cognizing has the structure and character of knowledge but may be and in interesting cases is inaccessible to consciousness’ (Chomsky, 1980, p. 188, cited from Rey, 2020, p. 269) Knowledge of grammar is of course not purely sensory, or affective, or emotional, to mention some domains of mind contrasted with the cognitive. Nor is it purely perceptual (its crucial role in speech perception notwithstanding), figuring also in speech production.

Could it be said knowledge of grammar is either a perception-like input system or a motor-like output system, in either case to be contrasted with cognition understood as a more ‘central’ resource or suite of resources? Aside from being suggestive of an oddly disjunctive, or anyhow disunified view of knowledge of grammar, not all manifestations thereof fit neatly in either category. They include processes of updating knowledge of grammar during acquisition. Even though these processes are, plausibly, highly innately constrained, there is reason to think they execute forms of probabilistic inference (cf. Lidz & Gagliardi, 2015, pp. 12–13). As Georges Rey (2020, pp. 176–179, 276–286) argues, drawing on this evidence, and following Leibniz, knowledge of grammar may be innate *and* learned. Another, and perhaps more telling, example is provided by processes implicated in delivering information on acceptability or interpretability. For example, such processes may yield the intuition that in the sentence ‘Bill claimed that the clerk deceived him’ the pronoun ‘him’ cannot refer to the clerk (cf. Dwyer & Pietroski, 1996). More broadly, such processes may have a role simply in thinking about how to express oneself clearly and effectively (Knowles, 2000, p. 332). In view of these various roles, James Higginbotham concludes:

[T]he rules of grammar, being manifested in behaviour and mental processes of most diverse kinds, are not at the service of any particular mechanism. They



are, in other words, a central resource, with many applications. (Higginbotham, 1987, p. 124, cited from Knowles, 2000, p. 332)

Some have argued, on these grounds, that knowledge of grammar should indeed be conceived as a body of belief (Dwyer & Pietroski, 1996) or propositional knowledge (Knowles, 2000). We need not, and perhaps should not, go so far however. Even if the indicated considerations only underwrite conceiving of knowledge of grammar in terms of somewhat more rudimentary forms of cognitive states and operations (for in-depth consideration of various options, see Rey, 2020), they still support conceiving it as a central cognitive resource.

Moreover, even if, despite these points, knowledge of grammar should not be classified as ‘cognitive’ in the precise sense relevant to Green’s proposed way of separating the perceptual from the cognitive, its status as dimensionally restricted would still bear on the comparative merits of his proposal versus one incorporating stimulus-control as necessary. Since the latter could explain the non-perceptual status of knowledge of grammar, it would, other things equal, have greater scope.

#### 4 Perceptually grounded demonstrative thought

While a function of stimulus-control is plausibly necessary to perceptual as opposed to cognitive status, I doubt it is sufficient. This section presents the first of three problems for that sufficiency claim, one noted by Beck and Phillips, viz. that of perceptually grounded demonstrative thoughts, e.g. the thought that *that* [heard] noise has *this* [heard] crackling texture.<sup>5</sup>

Such thoughts purport to refer to perceived objects or properties and to do so in a way exploiting their perceptual availability. The reference of the demonstrative components is determined not by a descriptive condition but by a perceptual relationship with the respective referents. A perceptual link, via incoming stimuli, is needed for the purported demonstrative reference to be successful (cf., e.g., Evans, 1982, p. 72, Campbell, 2002, pp. 8–9, Levine, 2010, pp. 177–178, Recanati, 2012, p. 62). These requirements of referential success arguably correspond to a functional requirement, upon the processes generating and maintaining these thoughts, that the thoughts be causally sustained or controlled by stimuli. Even so, demonstrative thoughts seem to be cognitive.

Beck’s response to this problem rests on three claims: (i) besides properly perceptual-demonstrative referential devices, demonstrative thoughts also need to include conceptual attributives among its representational elements, such as, in our example, the concepts *noise* and *crackling texture*, to guide successful demonstrative reference; (ii) such conceptual attributives can be used in ways that do not function to be stimulus-controlled; and (iii) demonstrative thoughts fail, therefore, a sharpened requirement on perceptual status, which Beck dubs (and adopts as his considered view):

---

<sup>5</sup> Henceforth, ‘perceptually grounded’ will generally be tacitly understood in front of ‘demonstrative thought’, except where otherwise clear.



**S-D FULL:**  $\Psi$  is perceptual if, necessarily, all occurrences of all elements of  $\Psi$  have the function of being stimulus-dependent; otherwise,  $\Psi$  is cognitive. (Beck, 2018, p. 330, boldface added)

By ‘elements’ Beck means representational elements, serving to represent some object, event, property, or kind.

Beck’s claim (i) here could be questioned. He suggests that without guidance from conceptual attributives, the purported demonstrative reference to an object or property would be left indeterminate; it would be indefinite what the (would-be) thought would be about. One might *mouth* ‘That is that’, pretending to express a purely demonstrative thought, but no determinate thought would be expressed. Beck does not consider here how attention might contribute to fixing demonstrative reference. Why could not object-based attention to certain salient object, *a*, and feature-based attention to a certain feature, being red, say, which object *a* looks to have, help to fix that a purported purely demonstrative thought that *that* is *thus* (as we might put it) has a determinate content to the effect that *a* is red?<sup>6</sup> Note that this suggestion is compatible with the idea, defended by Burge (2009, 2010), and invoked by Beck, that for object *a* to be perceptually presented in a way suitable for object-based attention and perceptual-demonstrative reference, some *perceptual* attributive, i.e. some nonconceptual representational capacity activated at the level of perception, must be deployed of *a*. Indeed, contrary to what Beck suggests, Burge’s (2009, p. 274) only argues that perceptual-demonstrative reference at the level of thought requires guidance from perceptual attributives, operating at the perceptual level, not necessarily from *conceptual* attributives in the thought.

It may be said that even if demonstrative thoughts *may* be purely demonstrative, in the sense indicated, they *need not* be. Therefore, if we let the type *demonstrative thought* be the type substituted for ‘ $\Psi$ ’ in S-D FULL, it will not be true that *necessarily* all occurrences of all elements of in a state of this type have the function of being stimulus-dependent. However, the type in question here, *demonstrative thought*, is individuated in terms of the kind of conceptual resources that, possibly in part, are deployed, viz. demonstrative concepts. It is, in effect, the type *at least partly demonstrative thought*. If that is an admissible instance of ‘ $\Psi$ ’, the type *purely demonstrative thought* should be one too. The latter type would be misclassified as perceptual under SD-FULL.

Even conceding Beck’s claim (i), his treatment of demonstrative thought encounters problems. First, a *prima facie* attraction of stimulus-control accounts, we saw, is that they permit perceptual representation to be partly conceptual or discursive, e.g. that a concept of car may be involved in seeing or visually classifying something as a car.<sup>7</sup> Second, and more importantly, a central motivation for stimulus-control accounts is the plausible idea that visualization and imagery may have the same type of representational contents and formats as perception, in which case some representational

<sup>6</sup> For a defence of such a suggestion, see Brewer (1999). Although that defence is couched within a conceptualist view of perceptual content, the claim here about the role of attention in fixing perceptual-demonstrative content is compatible with a nonconceptualist view, cf. Lerman (2012). See also Levine (2010, pp. 177–187).

<sup>7</sup> For an objection to S-D FULL based on how perception may involve use of concepts, see Quilty-Dunn (2020, p. 275).

elements active in vision may likely also be used in visual imagery. Now, if either the first or the second of the two last points hold, and if the possibility of deploying the conceptual attributives of demonstrative thought also in ways that do not function to be stimulus-controlled mean they fail to be perceptual under S-D FULL, then perceptions would be misclassified as cognitive by S-D FULL.

However, it may well be that Beck intends an interpretation of S-D FULL on which the the phrase ‘all occurrences ... have the function of being stimulus-dependent’ is restricted to occurrences *in states of type*  $\Psi$ . If so, the fact that a concept of cars occurrent in vision could also occur in, say, car memories, or a perceptual attributive of shape also in imagery, would be beside the point. Beck argues, plausibly, that concepts in (at least partly) demonstrative thoughts may function not so as to be stimulus-controlled *even* in those thoughts. For example, I can aptly judge that *that* bird [seen in the distance] is spotty, even when too far away for any spottiness to be visible, relying instead on information from memory (Beck, 2018, p. 330).

However, the concept of spottiness here precisely does not do the guiding job that Beck contends, in his claim (i) above, that non-demonstrative conceptual attributives are called upon to play in demonstrative thoughts. Even setting aside our doubts about his claim (i), it is *at most* for this guiding role conceptual attributives are claimed to be needed. The notion of guidance here, which Beck adopts from Burge (2009, pp. 275–289, 2010), has to do, broadly, with securing perceptual discrimination among various candidate referents. It is plausible that to serve this role the application of the relevant attributive must be causally sensitive to stimuli, for otherwise it is hard to see how it can play the role of selecting an appropriate informational link for the demonstrative element. Now, one type we may distinguish here, as an instance of ‘ $\Psi$ ’, is the type *at most conceptually guided demonstrative thought*, wherein any non-demonstrative conceptual attributives play at most this reference-guiding role. This type would be misclassified as perceptual under S-D FULL.

Another, independent concern about S-D FULL turns on representational devices accomplishing egocentric spatial or temporal representation in perception. Aside from representing putatively perceived objects or events, perceptual states also represent such dimensions as left and right. Even when nothing specific is perceived on the right of a light straight ahead, there is still a visual sense as of there being spatial regions on its right. Relatedly, perception arguably involves analogues of such pure indexicals as ‘I’, ‘here’, and ‘now’, what Burge (2009, pp. 256, 270) dubs ‘*de se* markers or egocentric indexes’. It is at least not clear that these representational elements function to be stimulus-controlled.

At this point, a friend of stimulus-control as sufficient for perceptual as opposed to cognitive status may be tempted to change tack. Perhaps demonstrative thought, or at least purely perceptual-demonstrative thought, should be considered perceptual after all, a conclusion Phillips comes close to affirming?<sup>8</sup> Alternatively, perhaps stimulus-control should be combined with a criterion in terms of representational format or

<sup>8</sup> Phillips (2019, p. 15) says perceptual-demonstrative thoughts are partly perceptual (in so far as they deploy perceptual-demonstrative elements) and partly cognitive (in so far as they deploy non-demonstrative conceptual attributives). He gives no reason for thinking there could not be purely demonstrative thoughts, where the attributives are perceptual-demonstrative concepts, and, in so far as they are, they would be correspondingly purely perceptual, on his account.

content, proposing that perceptual states function to be stimulus-controlled and, partly or wholly, have nonconceptual content or non-discursive format (cf. Block, 2023)? Our two next problems for the sufficiency claim will, inter alia, indicate that the moves lately considered would not be adequate.

## 5 Play-by-play announcing

Functions can arise not just from nature but via social institutions, e.g., from the tasks applying to the roles of such institutions. To carry out such a role will often be a skill. A skill can be here understood, broadly, as a sequence of activities, organised towards certain purposes or functions (cf. Fitts & Posner, 1967, pp. 1–2). Where a skill corresponds to an institution, these functions will often be defined by the relevant task.

Now, consider the following job description, from Field's *Career Opportunities in Radio*:

Position description. (...) [P]lay-by-play announcers ... are the eyes of the listeners during a sport event. They watch sporting events and report the actions they see to the listening audience on-air as the action is happening. Those who are successful colorfully describe the plays of a game so that listeners can actually visualize what is occurring minute by minute. (Field, 2004, p. 53)

A play-by-play announcer provides play-by-play announcements of the events on the pitch, as and when she sees them happening (henceforth, I skip 'play-by-play' before 'announcer' and cognates). When things go as they should, to a sequence of announcements,  $\langle A1, A2, \dots \rangle$ , there correspond to a sequence of impacts on the eyes of the announcer,  $\langle S1, S2, \dots \rangle$ , such that  $S1$  is a cause of  $A1$ ,  $S2$  a cause of  $A2$ , etc.<sup>9</sup> As such, when things go as they should, the announcements are stimulus-controlled. This assumes of course, that human actions, such as announcements (a form of speech acts), have causes, and that among their causes are not merely mental states such as beliefs and desires, but non-mental causes of such states, including retinal impacts. That assumption is however widely accepted.

Corresponding to the job description there is the skill of announcing. This is the sequence of activities or processes, organised to fulfil the task described, including that of ensuring that the sequence of announcements causally reflects impacts on one's eyes. By 'announcing' I mean (unless otherwise made clear) the process of carrying out this skill. Announcing subsumes sensory, perceptual, and motoric sub-processes. However, announcing also manifests conceptual, verbal, and broadly intellectual capacities in a way that makes it apt to class it as a cognitive process (as we shall return to in Sect. 6.2 below). Even those who reckon, with Millikan (2004, pp. 113–127), that the

<sup>9</sup> Various questions arise here. Where does one announcement end and another begin? Can the announcements in the sequence overlap, or be separated by gaps? These questions about individuation or sequencing arise also for stimuli—and indeed for perceptual states. Neither Phillips nor Stegmann (2014) make explicit any conditions of individuation or sequencing, as plausibly necessary for causal control, of a sort that would permit stimuli to control perceptions but *not* to control announcements. It is at least not obvious what conditions of that sort could be invoked. I shall therefore skate over these questions of individuation and sequencing, as they seem orthogonal to the issues at stake in this paper.

taking-in of testimony should be classed as broadly perceptual would not ipso facto class *producing* testimony as perceptual.

To be sure, announcers occasionally make remarks that are not stimulus-controlled, for example to advertise upcoming games, on statistics or the bios of players, etc. Some such remarks however are not made in their capacity as announcer, but qua also wearing the hat of, say, host or promoter of their station. Perhaps some announcers have it written into their role as announcers to occasionally interject some biography or statistics (lulls in the action permitting). However, they do not then perform the skill of announcing, in the strict sense relevant here. One way to bring this out is that there easily could be another school, call it ‘pure’ announcing, where nothing but reports on observable features of unfolding events, as and when they are seen to happen, is called for. Hereafter, ‘announcing’ is limited to skilfully carrying out such a pure form of the institution (except where otherwise clear from context).

Even a skilled announcer can occasionally make announcements that are not stimulus-controlled. She may hallucinate. She might get sand in her eyes, and, so as not to let her listeners suspect the mishap, make predictions, driven by theories or background knowledge, as to what is going on, and report accordingly. However, when such happens, she is not exercising the skill of announcing. Rather, she is covering-up impairments currently conflicting with that exercise. This is not an arbitrary restriction on what announcing, in the relevant sense, is all about. It can be supported in terms of how skilful announcing enables listeners to follow the game, as it happens. If listeners learnt that what they had been served were expressive of the theory-driven predictions of an unfortunate announcer whose retina temporarily got detached, then, even if the reports (stunningly!) were correct, they could rightly complain they had been led astray: they had not really been following the game. This suggests that it is a function, applying to the process of working in the capacity of announcer and carrying out that skill, to produce ‘representations of environment entities [viz. in the guise of announcements] by being causally controlled by those proximal stimuli that these entities produce’. To be sure, this is *a* but not *the only* function: the announcements are also supposed to be, say, veridical. The same goes for perception, however, as Beck (2018, p. 326) underscores: its having the function of being stimulus-controlled does not exclude its having other functions, such as that of veridicality.<sup>10</sup>

The school of pure announcing is, or could well be, a school of *visually* pure announcing. It could well be that announcements are to be responsive strictly to *visual* stimuli, i.e. to the optical impacts on the eyes of the announcer. Contrast ‘multimodal’ announcing, where announcement may be responsive to impacts on any sensory modality. These are different skills. Someone might excel at multimodal announcing but be a laggard at the visually pure form, e.g. because their other senses compensate for relatively poor eyesight. Under some circumstances, such as when non-visual cues tend to be treacherously misleading in the settings in which announcers operate, visually pure announcing may come to predominate, and be the in-demand skill. One fails

<sup>10</sup> It may be said it is a function of announcing that it be controlled (in part) by intentions and beliefs of the announcer. For our purposes, there is no need to deny that this may be so, for it would not follow that it is not (also) a function of announcing to be stimulus-controlled. In Sect. 6.1 below, I offer further support for the claim that the functions of announcing are not (merely) those of expressing the announcer’s beliefs.

at that skill if one's announcements are causally controlled not by what hits one's eyes but, say, auditory input. Since such visually pure announcing is still cognitive, it follows, *pace* Phillip's Stimuli-Specific-Control<sub>P/C</sub>, that not even a function to be causally controlled by modality specific stimuli is enough for perceptual as opposed to cognitive status.

## 6 Announcing: some rejoinders answered

This section answers some rejoinders to the problem announcing seems to pose for the sufficiency of a function to be stimulus-controlled for perceptual as opposed to cognitive status.

### 6.1 Not a function to be stimulus-controlled

A first rejoinder is that the functions of announcing all amount to something other than producing stimulus-controlled announcements. Now, as we saw in the last section, its function is not merely that of producing veridical announcements. Would it suffice to add functions of producing reports that are knowledgeable, or reliable, concerning what is happening on the pitch?

It would not. As a cover-up for her failing eyesight, an announcer may bribe players to play according to her detailed instructions. Safe in the knowledge of what will happen she, near blind, later makes knowledgeable, reliable pronouncements concerning the unfolding events but is hardly doing her job properly; she is not exercising the skill of announcing. Listeners fed such statements could rightly complain that they were not thereby enabled genuinely to follow the action as it happened.

How about specifying that the function is to generate announcements expressive of *perceptual* knowledge? This would have to mean, specifically, continually updated perceptual knowledge gained from perceiving events unfolding on the pitch; knowledge gained by perceiving, say, match-fixing preparations would not qualify. Now, for something to be expressive of this very specific type of perceptual knowledge is, arguably, perhaps in part, for it to be causally controlled by something, viz. continually updated states of knowledge, that in turn are causally controlled by incoming sensory stimulation. That is to say: for announcements to be expressive of such knowledge is, perhaps in part, for there to be sequences of stimuli,  $\langle S1, S2, \dots \rangle$ , states of knowledge  $\langle K1, K2, \dots \rangle$ , and announcements,  $\langle A1, A2, \dots \rangle$  so that  $S1$  causes  $K1$  which causes  $A1$ , and ditto for  $S2$ , etc. Assuming that the causal links here are such as to underwrite the transitivity of causation, this involves sequences such that  $S1$  causes  $A1$ ,  $S2$   $A2$ , etc., i.e., that stimuli causally control announcements. If this is so, a function to generate announcements expressive of the relevant sort of continually updated perceptual knowledge is a function to generate announcements that are causally controlled by states that are causally controlled by stimuli. It is a function the fulfilment of which necessarily consist in being causally controlled in a certain way by proximal stimuli. Thus, the indicated would-be alternative function comes at least very close to being, or entailing, a function to be stimulus-controlled.

Even if that would-be alternative does not quite entail the latter function, there is another obstacle to the claim that it is *no* function of announcing to be stimulus-controlled. It could, after all, easily just be written into the contract of announcers that their announcements be continually responsive to how the events on the pitch stimulate their very own eyes. Alternatively, it could easily come be tacitly agreed in the community of producers and consumers of announcing that announcing properly requires announcements to be stimulus-controlled. There could be intelligible, non-ad-hoc rationales for that functional requirement being treated as partially defining of skillful announcing. Beside the slightly outlandish worries about would-be announcing from detailed match-fixing plans, the noted requirement would ban a practice of one announcer relying on and in effect recycling the announcements of someone else. A function specification that appeals to what hits the eyes of the announcer might be easier to understand than a one appealing to a distinctive type of knowledge. A skill so defined may be easier to teach and nurture.

## 6.2 Announcing not a cognitive, psychological kind

A second rejoinder is that announcing, while it may be a process of a *social* kind, is not one of a *psychological* natural kind, whereas accounts of the perception/cognition-distinction, such as Beck's and Phillips's, should only aim to distinguish among psychological natural kinds. A related, perhaps more specific rejoinder is that announcing is neither cognitive nor perceptual, so belongs to neither of the categories between which such accounts purport to differentiate.

I reply that announcing is a skilled cognitive activity, meeting reasonable standards of natural kindhood for the purposes of psychology. Just what standard must be met to be a natural kind is highly contested (cf. Bird & Tobin, 2023). A moderate demand, assumed by Phillips, is that the members of a kind 'must share certain distinctive properties the appeal to which gives us scientific explanations and inductive inferences that we wouldn't otherwise have at our disposal' (Phillips, 2019, p. 317).

Skills can plausibly meet this demand. Psychologists have studied skills as diverse as telegraphy (William & Harter, 1899), flying (Fitts, 1947), bombing (Bartlett, 1947), batting (Schmidt & Lee, 2019), and haka (Mingon & Sutton, 2021), to mention just a few. These studies assume it is possible to make justified inductive inferences about the skills in question and give scientifically interesting explanations concerning their causes and effects. This research has had consequential applications. For instance, the work of Fitts, Bartlett, and co-workers was lauded for its contribution to aviation safety (cf. Jensen, 1986), and the allied war effort (Broadbent, 1997). The psychological study of skill is moreover vital field of ongoing research (cf., e.g., Schmidt & Lee, 2019). This indicates that the assumptions concerning inductive and explanatory integrity are borne out. If the noted, diverse skills are targets of apt inductions and explanations, why not also the skill of announcing?

Perhaps it will be objected that the social character of announcing means it cannot be a natural kind, on the ground that natural kinds are not socially or conventionally constituted. Now, although the tasks defining the skill of announcing have arisen as part of a social institution of announcing, it is not obvious that they couldn't have had

a different, individual source. An inventive individual could arguably design a system of announcing, supposed to deliver reports on observable events as and when they strike one's eyes, not for some interpersonal benefit, but because, say, reporting on those ongoing events may strengthen one's appreciation, understanding, or recollection thereof (as some people find it useful to think out loud about thorny problems). Be that as it may, the interpersonal character of announcing, such as it is, leaves it in company with plenty other skills, e.g. strategic skills such as dribbling, or artistic ones as haka. Beyond skills, such classic psychological kinds as selfhood (as studied in infant development, say) or jealousy are deeply interpersonal. Some 'moral' emotions may be social in other, perhaps less obvious ways; for example, there may be grounds for distinguishing shame as it exists in the Western societies from shame as it exists in east Asian as two kinds of emotion (cf., e.g., Mesquita, 2022). This suggests that the kinds that are of inductive and explanatory importance to psychology may well be social in deep and interesting ways. Either psychology should not be restricted to natural kinds, or, more plausibly perhaps, its natural kinds not to the non-social.

The voluntary character of announcing does not prevent it from being cognitive. Several cognitive processes, e.g. reasoning from a counterfactual supposition, are voluntary activities. Its status as verbal activity is not a block to being cognitive. Quite the contrary: psychologists often class skills as higher-level, intellectual skills, contrasting with perceptual-motor skills, precisely on the ground on being verbal (cf., e.g., Fitts & Posner, 1967; Rosenbaum et al., 2001). In philosophy, thinkers as different as Plato (2014, p. 189e) and Hobbes (1996, p. I.iv) agree in taking meaningful speech to be a model of or even partially constitutive of thinking—in particular: of the general, rational thinking of which humans are capable. More recently, so diverse theorists as Brandom (1994, p. xv) and Williamson (2000, pp. 255–256) treat such paradigm cognitive acts as judgement as analogues of assertion. The cognitive character of verbal processes is buttressed by various experimental findings. For example, people often need to 'think in words' to solve certain reasoning tasks (that do not themselves concern language). If their verbal capacities are drained by another, irrelevant verbal task (shadowing a stream of speech), performance suffers, whereas being encumbered with an equally taxing non-verbal task has much less of an effect (Herner-Vasquez et al., 1999). From the perspective of higher-level psychological-cum-philosophical theories of mind, there are reasons for holding conscious, reflective, 'System 2'-type thinking characteristically to be linguistic in form (cf. Frankish, 2018).

It might be objected that what we are calling announcing is just a collection of perceptual, cognitive, motoric etc. sub-processes; that neither of these sub-processes are both cognitive *and* function to be stimulus-controlled; and that accounts of the perception/cognition-distinction should limit attention to these sub-processes. However, even paradigm perceptual and cognitive processes are rife with sub-processes: shape perception, e.g., involves edge detection, mental arithmetic includes working memory, etc. Moreover, it is implausible to claim that announcing is a merely collection sub-processes. In general, one of the hallmarks of skill is precisely the appropriate timing and coordination of component sub-processes (Rosenbaum et al., 2001, p. 464), and this is eminently so for announcing.



### 6.3 The function merely contingent

A third rejoinder draws on a point from Beck (2018, pp. 324–325). He envisages a brain-manipulating helmet, which ensures that, when a certain proximal stimulus is imposed, a belief that Omaha has 434,353 inhabitants is triggered, and causally sustained until stimulus offset. Although the token belief here is, in a sense, stimulus-dependent, this does not reflect any necessary feature of the kind of state—belief, or belief with such-and-such content—to which it belongs. However, the sort of stimulus-dependence that distinguishes perceptions from cognitions is, Beck argues, supposed to be one that necessarily applies to them qua the type of mental state they are. Now, inspired by this sort of point, it might be objected that announcements are assertions, and that it is not true that, necessarily, assertions are stimulus-controlled or have the function of being.

However, for perceptions, as for announcements, one can ask about the relevant kind for assessing the contingency claim at stake here. If (as is currently assumed) perceptions are representational, they belong at least to the following kinds: mental states; representational states; representational states with a mind-to-world direction of fit. It is not true that, necessarily, instances of one or another of these three types have the function of being stimulus-controlled. As Beck (2018, pp. 326–327) later argues, a function of being stimulus-controlled, in so far as it applies to perception, plausibly is to be explained in terms of a function, viz. a function of generating such-and-such states, characterizing the producing mechanisms or processes. Therefore, if the function of the state is to necessarily apply to the kind of state it is, that kind must likewise reflect something about the producing mechanism. The situation for announcements here very much seems analogous to that for perception. Qua mental act, representational act, and representational act with affirmative force, announcements are not necessarily stimulus-controlled, nor function to be. But qua the sort of mental act that reflects their specific generating process, viz. the skilled activity of announcing, that function necessarily applies.

That the process of announcing (and, so, the announcements generated) necessarily is subject to such a function does seem plausible. Suppose an announcer is handed a bunch of scripts for upcoming games by shady higher-ups, who insist that she will, from now on, be reading out on air from the contents of these scripts, reassuring her that her reports (or ‘reports’) will be accurate to the facts. She could rightly protest: that’s an utterly different job—it does not involve the skill of announcing! It is not as if they had merely asked her (not) to use received pronunciation.

### 6.4 Insufficiently tight stimulus-control

A fourth class of rejoinders complains that even if announcing functions to be stimulus-controlled in some loose sense, the form of control in question is not as tight as we should demand of perception, and as stimulus-control theorists intend.

One way to press this charge is by appeal to Beck’s ‘S-D FULL’, discussed in Sect. 4 above, to repeat:

S-D FULL:  $\Psi$  is perceptual if, necessarily, all occurrences of all elements of  $\Psi$  have the function of being stimulus-dependent; otherwise,  $\Psi$  is cognitive. (Beck, 2018, p. 330)

As we saw, for this not to misclassify perceptions as cognitive, the phrase ‘all occurrences of all elements of  $\Psi$  have the function of being stimulus-dependent’ should be interpreted as restricted to occurrences in state of type  $\Psi$ . As noted, a worry arises here concerning elements of perceptual states accomplishing egocentric spatial or temporal representation. Setting this aside, it is far from obvious that announcements fail to class as perceptual under S-D FULL, on the relevant interpretation. ‘Elements’, recall, allude to devices representing some object, event, property, or the like. In (pure) announcing, one should only report on observable features of seen events, as and when and because they strike one’s eyes. Any reference to a player or event, or attribution of a property, should be under the control of stimuli carrying information of such things. It is true that there may be no systematic mapping between types of optical stimulation and types of reports on the ongoing events. As Beck stresses however, such a mapping cannot be required for perception, since a great variety of local proximal stimuli may yield a representation of a constant distal property (e.g. in shape or colour constancy) and a constant local stimulus may yield diverse representations of distal properties, for a variety of reasons including context effects and, arguably, cognitive penetration. Beck’s requirement, then, is only that a perceptual representation ‘be causally sustained by *some present proximal stimulation or other*’ (2018, p. 331, Beck’s italics).

Another way of pressing, or buttressing, the charge that announcing is marked by a ‘lack of tightness’ in stimulus-control is as follows. Even in (pure) announcing it is, surely, okay to remark on players that are briefly hidden from view, e.g. by other players tacking them. Hidden objects are not seen, and no stimuli are received therefrom. This suggests a difference in tightness in stimulus-dependence between the perception of objects, on the one hand, and the announcing thereon, on the other, or so the objection goes.

Now, there are various cases to consider here. Sometimes covered objects are seen, as when a mouse runs about under a thin blanket on the floor. Some announcing on covered players may be akin to such situations. It is also important to keep in mind here that stimulus-control accounts of the current stripe do not purport to mark out *perception of objects*, understood in the success sense, but *perceptual representation as of so-and-so objects*, a class that, by supposition, includes some hallucinations, where no object is perceived.

It is plausible that there may be perceptual representation as of objects, or parts of objects, that are occluded (and not because of hallucination or ‘seeing mouse running under the blanket’-type cases). At least some forms of amodal completion, as when a dachshund is seen to hang evenly together behind a flagpole, and the tunnel effect, where an object is seen to pass continuously behind an occulder, illustrate this.<sup>11</sup> This is consistent with stimulus-control, suitably construed, since such completion is sensitive to details of the broader proximal stimulus, e.g. concerning occluding

<sup>11</sup> For the claim that some amodal completion is perceptual, see Kanisza and Gerbino (1982). For the tunnel effect, see, e.g. Flombaum and Scholl (2006).

edges, the features of the (partially) occluded object, etc. Still, the stimulus-control requirements, whatever they are, must be such as to allow for perceptual representation as of occluded parts of objects or briefly occluded objects. So, it is unclear why such requirements would be unsuitable for announcing, even if announcing permits representations of briefly occluded players or occluded parts of players. Perhaps it will be said that it must, surely, be admissible to go much further in describing occluded action in announcing than perception can go in representing the occluded. Yet why should not announcers who go far in that direction be faulted for embellishing their art with elements quite foreign to the true skill of pure announcing?

## 7 Announcement-driven visualizing

This section introduces the third problem for the sufficiency of a function of stimulus-control to perceptual as opposed to cognitive status. This process, which I dub ‘announcement-driven visualizing’ (ADV), differs interestingly from the two foregoing problems in having a more vision-like output.

Theodore, recently blind, misses watching a game, and sets out to design a way in which people like him can follow a game enjoyably. To follow a game, on his (not implausible) understanding, requires entertaining representations of so-and-so going on just now, because so-and-so indeed is going on just now (where ‘just now’ has enough flexibility to make room for some delay in information transmission). Realising that he has a talent for voluntarily generating visual imagery, that entertaining imagery of the action is enjoyable (whereas say ‘thinking in words’ about it is boring), and that others can visually take in and narrate events to him as they occur, Theodore devises a system with the following sub-processes. First there is, in effect, announcing (not yet a familiar institution in his time, we may imagine). Second, there is process taking certain auditory stimuli, corresponding to the announcements, as input and yielding imagery as output, via voluntary generation, so that imagery is continually updated in consequence of incoming auditory stimuli. When all goes well, implementations of this system allow people with Theodore’s talent for imagery to follow a game enjoyably. It enables them to enjoyably see the game in their mind’s eye, as one says, without literally perceiving it.

Within the system, the functions of the sub-process of announcing are pretty much as described in preceding sections. The second sub-process distinguished, which I label ‘announcing-driven visualizing’ (ADV), has a function of taking auditory stimuli, caused (albeit indirectly) by the action on the pitch, as input and, in consequence thereof, generating imagery of the action, in a manner ensuring that auditory stimuli causally control the imagery. This claim about function follows given the widely accepted claim that something has the function F if it was designed to F (cf. Kitcher, 1993, and, for further references, Nanay, 2010, p. nt. 35).<sup>12</sup> The claim can also be supported in terms of a Cummins-style functional analysis of Theodore’s system, i.e. of how it enables enjoyable following of a game (Cummins, 1975). The claim can

<sup>12</sup> Processes may be designed no less than tangible tools and machines. Process design is indeed central to many branches of engineering, e.g. logistics engineering.

moreover be supported in terms of the idea that X can have the function F in virtue of the fact that using or implementing X would, by X's F-ing, contribute to fulfilling the implementor's goals (cf. Nanay, 2010, pp. 428–431). For, by ensuring imagery is controlled by auditory stimuli, ADV-ing would contribute to the goal of enjoyably following the game. A variety of views of functions, then, underwrite the idea that ADV has a function of generating stimulus-controlled imagery.<sup>13</sup>

There is no claim here, it might be worth stressing, that all who see the game in their mind's eye, hearing announcements thereof, engage in ADV. If, for example, the imagery is generated involuntarily, as intrusive imagery, caused by overhearing certain announcements, there is no ADV-ing. The suggestion is merely that ADV, with the indicated functions, is a nomologically possible process.

This amounts to a challenge to the sufficiency of a function of stimulus-control for perceptual as opposed to cognitive status that differs interestingly from those of announcing and perceptual-demonstrative thought. Its output, visual imagery, is explicitly considered cognitive in the stimulus-control approach. Still, it is phenomenologically and representationally akin to vision. In Sect. 3 above, we outlined the case for thinking it has non-conceptual content, or non-discursive format, assuming vision does. ADV thus indicates that perception cannot be distinguished from cognition even by taking the former to combine a function of stimulus-control with a non-conceptual content or non-discursive format.

## 8 ADV: some rejoinders answered

Counterparts to the rejoinders concerning announcing could be raised for ADV; broadly similar replies would apply. Below, I address two rejoinders trading on distinctive features of ADV.

### 8.1 Is ADV akin to sensory substitution systems, and so perceptual after all?

The first objection runs so: 'Even if imagery ordinarily is to be classed as cognitive there are specific reasons to consider the imagery—or, to introduce a neutral term: the intrinsically perception-like states—outputted in ADV perceptual, viz. from the analogy between ADV and use of sensory substitution devices (SSDs). These devices capture optical information with a camera worn or carried by the user. Optical information is converted to tactile input on the user's back or tongue, or auditory input in headphones. Blind or blindfolded users of SSDs achieve intrinsically perception-like states—sometimes described as cases of visualizing (Renier et al., 2005) or imagery (Nanay, 2017) – of such camera-captured features of surrounding objects as their locations, overlap, and even shading (cf. Pence, 2021). It is widely held, moreover, that when all goes well subjects indeed *perceive* the relevant distal features, either visually (cf., e.g., Pence, 2021; Renier et al., 2005) or in some other, say tactile or 'metamodal' way (cf., e.g., Martin & Le Corre, 2015). ADV would, then, seem to be

<sup>13</sup> Points made in the last two paragraphs could be marshalled also to bolster the function ascriptions to announcing in Sects. 5 and 6, if some such bolstering were desired.

analogous to SSD-use: both involve on-the-fly generation of intrinsically perception-like states causally guided and controlled by stimuli, albeit, in each case, in a somewhat non-standard way. So why not say that ADV-ers likewise perceive the action?'

Now, certainly not all SSD-use amounts to perception of the distal, camera-captured features of the surrounding. Early in training with the devices, users are typically acutely aware of the properly tactile or auditory input. On the basis of this input and an open-ended set of information, including that provided explicitly during training concerning how the devices work, and background knowledge concerning various categories of objects, users infer what distal features of the surroundings are indicated (cf. Deroy & Auvray, 2012; Siegle & Warren, 2010).

It is true that sometimes, often after extensive training, there may be a remarkable shift. Users' experience has been reported to change so that they are no longer aware of stereotypically tactile/auditory information delivered in the 'substituting' modality. For example, with visual-to-tactile devices, the input has been said to be felt as no longer upon the skin but located in some distal location, corresponding to the distal properties (allegedly) perceived (Guarniero, 1974). The processing by which these so-called 'distal attributions' are formed also become automatic or involuntary, and bottom-up (cf. Nanay, 2017; Pence, 2021).

Even in these cases there is debate however whether perception is achieved. Ophelia Deroy and Malika Auvray (2012) argue that even here the process is more akin to reading and comprehending a text, corresponding to the tactile/auditory input, than to perception. If they are right, the argument by analogy that ADV is perceptual is of course undercut. If they are wrong, that argument is still hampered, because it would underscore that the analogy is quite weak. Aside from the noted contrast that 'distal attribution' in well-adapted SSD-use is involuntary whereas ADV involves voluntary mental imagery, imagery in ADV is formed on the basis of comprehending linguistic input, forming beliefs about what the announcer is conveying, whereas, if Deroy and Auvray are wrong, there is no such comprehension-like process in well-adapted SSD-use.

In the next section will argue that processes of belief fixation, comprehension, and voluntary imagery have a certain non-modular character, and that the involvement of such processes in ADV accounts for its non-perceptual status.

## 8.2 Visualising requires creative enrichment, so ADV cannot function to be stimulus-controlled?

The second rejoinder runs as follows: 'In visualizing the action, one inevitably visualizes it in a way that fleshes it out in certain respects, e.g. concerning its setting (as taking place on a green, flat surface), the rough shape of the participants (typically humanlike), or the like. This is so even if the announcer says nothing about such taken-for-granted aspects of the action, as she typically would not. These aspects of what one is visualizing would not, then, be stimulus-controlled. Moreover, their failure to be stimulus-controlled cannot plausibly be construed as a case of malfunctioning, for some such 'creative enrichment' is inevitable in imagery. After all, visualizations are supposed to be minimally vision-like, in some sense. Now, one cannot genuinely *see*

action on the pitch without having at least some vague, generic impression concerning the ground on which it unfolds, the shape of the players, or the like.<sup>14</sup> Likewise, one cannot really visualize it without some imagery, if only vague or generic, concerning such features.’

It true that an announcement that explicitly describes merely a certain football move underdetermines subsequent imagery that *inter alia* imagines it as happening on a green backdrop. Yet given, as a background causal condition, the hearer’s belief that football likely unfolds on green surfaces, the announcement may yet cause that imagery in the sense needed for causal control. Our paradigms of causal control also presume certain background conditions. The studs on the music roll causally control tones only given the presence of air. The base pairs in DNA causally control amino acid concatenation only given the biochemical environment in the cell.

Of course, perception itself is causally controlled by stimuli only given suitable background conditions. One important reason why this is so is that stimuli underdetermine perceptions, being ambiguous as to what distal situation they signal (cf., e.g., Palmer, 1999; Rock, 1983). It is widely assumed, therefore, that stimuli causally control distal representations only given certain implicit assumptions or Bayesian priors, such as, e.g., the ‘light from overhead’ prior (cf. Hershberger, 1970), or ‘slow motion’ prior (cf. Weiss et al., 2002). At one level of abstraction, there is a commonality between how verbal stimuli cause imagery of green backdrops, via background beliefs about the colour of pitches, and how retinal impacts cause perceptual representation of, say, convexity or slow motion, via assumptions of the indicated sorts. This is not to deny, of course, that are also differences; the challenge is to account for what they are.

Could those who take a function of stimulus-control to suffice for perceptual status argue that the noted assumptions or priors in visual processing, unlike the background beliefs at work in voluntary generation of imagery, themselves function to be stimulus-controlled? However, the noted assumptions or priors are widely considered to be innate (cf. Hershberger, 1970; Scholl, 2005). Even if they may undergo updating under the course of development (cf. Scholl, 2005), that does not underwrite a function to be stimulus-controlled; by that standard, plenty of high-level empirical beliefs, acquired and updated in response to stimuli, would also qualify as having that function.<sup>15</sup>

Alternatively, it could be argued that the background beliefs at work in voluntary imagery are explicit or ‘psychologically real’ in various senses in which the visual assumptions or priors are not; the latter are merely implicit in the processing (cf., e.g., Block, 2018). Even if that is so, however, it does not explain why the former as opposed to the latter could not mediate in processes of stimulus-control. Of course, one might invoke a notion of (non-)explicitness or (lack of) psychological reality to

<sup>14</sup> Cf., e.g., Pautz (2021, p. 131) on ‘No radical incompleteness’.

<sup>15</sup> Perhaps it will be said that, even if these priors, as standing characteristics of the visual system, are not stimulus-dependent their *activation* is stimulus-dependent, for they only come into play when the visual system is stirred to action by stimuli. However, by the same token, the activation of background beliefs in ADV or announcing is also, when all goes well, the result of stimuli. It may be said that these beliefs, unlike the priors of the visual system, may also be activated by such internal psychological factors as intentions to ponder certain questions. Yet this may, in a sense, go for the priors in the visual system too. For part of the visual system, and so at least some of the priors implicit in its processing, is active also in the generation of visual imagery (cf., e.g., Dijkstra et al., 2019).

pick out a certain variety of stimulus-control held to be characteristic of perception. I am sympathetic to this sort of move. Notice, however, that it shifts towards a view on which broadly architectural or ‘mode-of-processing’ notions are needed to constrain a certain *subvariety* of stimulus-control characteristic of perception. Moreover, as such stimulus-control theorists as Beck (2018, p. 322) and Phillips (2019, p. 331) argue, stimulus-control allows for cognitive penetration, i.e. that high-level beliefs may modulate the processing leading from stimuli to perceptual states. Again, then, a causally modulating role for general background beliefs in voluntary imagery is no block to stimulus-control. Since we arguably should allow for some cognitive penetration, the broadly architectural notions in terms of which ‘properly perceptual’ stimulus-control is to be characterized need to allow the possibility of (explicit, real) beliefs playing a modulating role. The next section considers what notions may serve this end.

## 9 Modular stimulus-control

Why are our three problem cases—demonstrative thought, announcing, and ADV—cognitive rather than perceptual? Not, we have argued, because they fail to function to be stimulus-controlled. Nor because they in each case differ in representational format or content from perception, as we saw with ADV.

In each of the three cases, the outputs arise from stimuli via processes rich in psychological mediation. In this section, I first reject a broader diagnosis of how this might make them non-perceptual, pointing to personal-level stimulus-mediating states. I then propose a certain lack of modularity as a better explanation.

### 9.1 No personal-level stimulus-mediating states in perception?

Perceptual-demonstrative thought, announcing, and ADV all involve what may be dubbed ‘psychological stimulus-mediating states’, i.e. psychological states, *M*, such that, when the function of stimulus-control is fulfilled, it is fulfilled because stimuli cause *M*-states that in turn cause the relevant perceptual-demonstrative thoughts/announcements/announcement-driven imagery. In perceptual-demonstrative thought, the *M*-states are or include perceptions; in announcing, beliefs on the part of the announcer; in ADV, beliefs and states of utterance comprehension on the part of the visualizer. Moreover, the stimulus-mediating psychological states here are in each case personal level: they are conscious or available to consciousness, amenable for inferential integration with the subject’s central stock of beliefs, and available for guidance of a diverse range of purposive actions within the agent’s repertoire. While perception may allow *sub*-personal stimulus-mediating states (along the lines of, say, Marr’s, 1982 primal sketch), it may be suggested there is no place for personal-level stimulus-mediating states in perception.<sup>16</sup>

<sup>16</sup> This account of why perceptual-demonstrative thought, announcing and its ilk are non-perceptual has often been put to me in discussion. (Notice, incidentally, that this account permits cognitive penetration of



This account is too restrictive, however, for one perception may casually depend on another perception. A perception of overlap may cause a perception of difference in depth (cf. Palmer, 1999, pp. 236–237). A perception of edges and surfaces may cause perception as of a solid object (cf. Burge, 2010, p. 345). A perception of proximity may cause a perception of grouping (cf. Rock, 1983, pp. 75–76). A perception of form, motion, and configural relations among two objects may cause a perception of one as chasing the other (cf. Gao & Scholl, 2011). Such cases could be multiplied. In at least some of them, the causing perception has the hallmarks of personal-level status: it is conscious, and its content is available for inferential integration with the agent’s wider cognitive economy and to guide a diverse range of actions (cf. Rock, 1983, pp. 283–299).

## 9.2 In perception, personal-level stimulus-mediating states must be inputs and outputs of broadly modular processes.

What differs, then, between a case where one perception causally depends on another perception and a case where a cognitive state or act depends on a perception? Specifically, what is the difference, given that is not, or not only, a matter of whether the process that generates the output functions to be stimulus-controlled, or of whether the output has a non-discursive format or non-propositional content? It is hard to see a good alternative here to looking more closely at the character of the processes that generate, respectively, the perception from a perception and the cognition from a perception. In each case, the process in question takes as input a (personal-level) perception, one that, qua personal-level, is available to and exploitable within the agent’s economy of beliefs and intentions. However, when drawn on by a perception-generating process, the process ensures a significant isolation from these beliefs, intentions, and other attitudes that constitute the overall perspective or outlook of the agent. In contrast, when the perception enters into a cognition-generating process, the process secures a significant integration with that overall perspective or outlook. Loosely speaking, therefore, the contrasting character of the relevant processes ensure that the output representation remains ‘the work of the senses’, in the perceptual case, whereas, in the cognitive case, that output can be seen as a case of ‘what the agent makes of what her senses put to her’. That, anyhow, is the picture I want to suggest.

Less picturesquely, the proposal is that perception-to-perception processes are modular. Now, there are a variety of notions of modularity on the market. A comparatively minimalist proposal, due to Burnston and Cohen (2015), is to understand modularity in terms of a lack of what Fodor (1983) terms isotropy. To borrow a formulation from Green: ‘To say that a psychological process is isotropic is to say that, in principle, it has access to any of one’s beliefs, desires, intentions, and so on during normal functioning.’ (Green, 2020, p. 338) Non-isotropic processes, in contrast, would be significantly limited in what parameters they can take into account. Famously, Fodor argued belief fixation is isotropic in this sense, that its normal function allows one to, say, take one’s botany to bear on one’s astronomy. Mere non-isotropy makes for a comparatively

---

Footnote 16 continued

perception by (personal-level) belief, since the penetrating beliefs will not typically be caused by the very stimuli that cause (directly or indirectly) the penetrated perceptions.).

weak notion of modularity.<sup>17</sup> Further requirements on modularity might be imposed, e.g., Green's notion of dimensional restrictedness (cf. Sect. 3 above), which would imply, but is not implied by, non-isotropy. For present purposes I assume modularity requires at least non-isotropy, leaving open whether additionally to impose such further requirements as dimensional restrictedness.

I propose that perceptual processes function to be stimulus-controlled and meet a fairly minimal requirement of modularity, in the following sense:

**Modular Stimulus-Control<sub>P/C</sub>** A process generating psychological states, E, as of environment entities is perceptual only if

- (i) it has the function of producing the Es by being causally controlled by proximal stimuli that these entities produce, and
- (ii) for any personal-level stimulus-mediating states, M, in the process (i.e. states such that the function of stimulus-control is fulfilled because stimuli cause Ms and Ms cause Es), the sub-processes generating Ms and Es from Ms are modular.

Condition (i) here is the right-hand side of Phillips's Stimulus-Control<sub>P/C</sub>. Condition (ii) can hold because (a) no psychological states mediate in generating the output states, E, from stimuli; or because (b) at most subpersonal states mediate; or because (c) some personal-level states mediate, but these all figure in modular processes. Would the process generating the outputs be modular even in cases (a) and (b)? If modularity requires merely that the process is not isotropic, this is plausibly so, whereas, if modularity requires dimensional restrictedness, the matter is less clear.<sup>18</sup> Now, I leave open this question concerning precisely what modularity requires beyond non-isotropy. So I leave open whether (ii) could be replaced by the simpler:

(ii\*) the process is modular (as are its sub-processes).

Modular Stimulus-Control<sub>P/C</sub> only imposes, then, a moderate, or even conditional, requirement of modularity: i.e., at least not isotropy. Whether we adopt condition (ii\*) or the possibly somewhat weaker condition (ii), we can however explain why perceptual-demonstrative thought, announcing, and ADV are non-perceptual, as I shall now argue.

### 9.3 The non-modularity of belief fixation, comprehension, and voluntary imagery

There are powerful reasons to think belief fixation is isotropic (cf., e.g., Fodor, 1983; Samuels, 2006, 2012). This extends to the processes generating purely perceptual-demonstrative thoughts, of the form *That* [referring to a perceived object] *is thus*

<sup>17</sup> There could however be even weaker notions of modularity. For example, one defined in terms of Carruthers's (2006) notion of 'wide-scope encapsulation' would arguably not imply non-isotropy, cf. nt. 20.

<sup>18</sup> If a psychological state figures in an isotropic process, it figures in a process that can access the subject's beliefs, intentions, etc. One might expect, then, that it would be inferentially integrated with these various beliefs, capable serving to guide various purposive actions, and as such be personal-level. However, if modularity in the relevant sense requires dimensional restrictedness, it is less clear why processes that are non-modular because dimensionally non-restricted (i.e. cognition can add to the dimensions they can take into account in their computations), could not draw on subpersonal states.

[attributing a property the object is perceived as having]. There is an open-ended range of ways in which one's belief set may contain, or easily could come to contain, considerations prompting one to reject that thought in favour of, say: *That is not thus (it only looks to be thus)*. They include, for any proposition,  $p$ , such that you believe *If that is thus,  $p$* , consideration that could prompt you to think *Not  $p$* . This does not mean that the process actually accesses each and every belief—that it executes an exhaustive search of memory—only that, in its normal operation, it can access any one out of an open-ended variety of beliefs.<sup>19</sup> If this is so, perceptual-demonstrative thought fails condition (ii) of Modular Stimulus-Control since stimulus-control (and so successful reference) is fulfilled here because stimuli cause perceptions that cause the thoughts, but, at the latter step, by a non-modular process. Since beliefs figure as stimulus-mediators in announcing and ADV they would fail this condition too.

More specifically, beliefs in ADV depend on comprehending what is conveyed or said by the announcement-utterances hitting one's ears. There are reasons to think such comprehension is isotropic, e.g. in view of the open-ended character of the contextual information that can be brought to bear. Stanley observes:

When someone utters the sentence 'The policeman arrested the robber. He was wearing a mask', we generally interpret the pronoun 'he' as referring to the robber, rather than the policeman. We arrive at this interpretation by exploiting inferences about the plausibility of interpreting the pronoun in different ways, inferences guided by our knowledge of meaning together with background knowledge about the world. Virtually every sentence we hear contains context-dependent expressions. (Stanley 2005, pp. 1–2)

The open-endedness of the background knowledge that can be drawn on is well illustrated by this variant from Allott (2019, citing, and adding a twist to, Recanati, 2004, pp. 31–32): in the sentence 'A policeman arrested John yesterday; he had just taken a bribe' the reference assigned to 'he' may depend on beliefs about corruption among local police. This is not to deny that comprehending what is said may involve modular sub-processes, e.g. parsing, or assignment of minimal propositions (cf., e.g., Borg, 2004), or that is fast and often automatic, only that it is entirely modular, in a sense that excludes isotropy.<sup>20</sup>

Again, voluntary imagery seems to be isotropic. Since voluntarily generating imagery can be, an often is, a rational activity it seems it the process should have access to the agent's intentions. An open-ended variety of beliefs and suppositions can also inform the imagery one voluntarily generates. If one thinks, say, that astro turf is turquoise, and is led to believe the announced action happens on astro turf, one may well and perhaps likely will imagine the action as unfolding on a turquoise surface.

<sup>19</sup> For this point, and for compelling criticism of an exclusively modular view of central cognition and of Fodorian sceptical conclusions about the impossibility of cognitive science motivated by isotropy, see Samuels (2012, pp. 72–74).

<sup>20</sup> Some claims that comprehension is entirely modular, e.g. Sperber's, are allied to a massively modular view of mind; for a critique of such, see, e.g., Samuels (2012), who also points out that some weak notions of modularity or encapsulation, such as the 'wide-scope encapsulation' of Carruthers (2006) is consistent with isotropy, given that isotropy does not require exhaustive search. For a proposal as to how comprehension can be fast, and often automatic, yet unencapsulated, see Allott (2020).

If one supposes tango dancers are tactful and tactful people smile, one may well, and perhaps likely will, imagine tangoing people as smiling. This not to say that any given belief or supposition,  $p$ , could influence voluntary imagery, as it were in isolation. However, for virtually any such belief or supposition, there is a set of other possible beliefs or suppositions implying that, if  $p$ , then so-and-so would look this way rather than that, in combination with which it easily could affect how one imagines so-and-so.

These remarks do not of course purport to be a full-dress defence of the isotropy of belief fixation, comprehension, and voluntary imagery, projects well beyond the scope this paper. They are indicated merely to suggest its plausibility.<sup>21</sup>

#### 9.4 Some final remarks

Modular Stimulus-Control<sub>P/C</sub> imposes a rather modest demand of modularity, e.g. in that it permits cognitive penetration. Even if modularity requires also dimensional restrictedness (i.e. cognition cannot add to the dimensions perceptual processes can take into account), cognition could still, as noted, affect what *value* perceptual processes output on a given dimension, such as orientation or hue (cf. Green, 2020, pp. 371–381).

Is modular stimulus-control, in the present sense, also enough for perceptual as opposed to cognitive status? I am unaware of any case suggesting it is not, and so tempted to hypothesise that it is. I do not purport to have shown, though, that perception need not, even in part, be non-conceptual in content or non-discursive in format. Nor do I purport to have shown that it need not have a certain characteristically perceptual phenomenology. One question of interest here is whether modular stimulus-controlled processes are bound to output such representations/phenomenology. If so, modular stimulus-control would, in any case, at least nomologically suffice for perception-as-opposed-to-cognition. If not, we must grapple with the relevant cases of modular stimulus-control without perception-like representation/phenomenology: are they perceptual, cognitive, or perhaps irresolvably borderline? These are however questions for another occasion.<sup>22</sup> The contention here is that Modular Stimulus-Control<sub>P/C</sub> explains why a diverse range of phenomena are not perceptual, and that, failing a more appealing alternative, this gives some abductive support in its favour.<sup>23</sup>

<sup>21</sup> Imposing requirements beyond non-isotropy on modularity facilitates defending the non-modularity of belief fixation, comprehension, etc., but of course raises the bar for modularity of perception. For a case that perception is modular, by a standard of not only non-isotropy but dimensional restrictedness, see Green (2020).

<sup>22</sup> Another question in this neighbourhood is whether (non-)isotropy is a matter of degree, or could, or should, be reconstrued as such (the same could be asked concerning stimulus-control). I see no reason why it should be impossible to introduce degreed counterparts or analogues of these notions. Whether it would be preferable to (re-)cast Modular Stimulus-Control<sub>P/C</sub> in such terms depends on the extent to which it would allow for a more illuminating or explanatorily useful ordering of various psychological phenomena (including low- to higher-level perception, core cognition, or what have you). This question falls, though, beyond the scope of this paper.

<sup>23</sup> This argument for perceptual modularity supplements more familiar arguments, e.g. from the persistence of known illusions, or evolutionary considerations.

## 10 Conclusion

Rock, affirming various deep commonalities between perception and thought—that they both involve broadly propositional representation, and inference, often based on (consciously accessible) perception –, pointed to the following major differences:

[P]erception differs from thought primarily because it is rooted in and constrained by the necessity of accounting for the proximal stimulus. (...)

The other major difference between perception and thought is that perception is based on a rather narrow range of internalized knowledge (...) Perception must rigidly adhere to the appropriate internalized rules so that it often seems unintelligent and inflexible in its imperviousness to other kinds of knowledge. (Rock, 1983, pp. 339–340)

The view of this paper has affinities with this view of Rock's here. Perception is sensory. The idea that it functions to be causally sustained or controlled by stimuli brings out an aspect of its sensory status. Having a function to be stimulus-controlled is not however enough to be perceptual as opposed to cognitive, as we have seen in the diverse cases of perceptual-demonstrative thought, play-by-play announcing, and announcement-driven imagery. Their failure to be perceptual cannot, in each case, be explained by an absence of perception-style representation, or a presence of personal-level psychological states causally mediating between stimuli and outputs. A better explanation invokes non-modular processing. Their reliance on personal-level states, in a way that integrates with the attitudes that constitutes the agent's overall outlook, make their outputs apt to be regarded as cases of 'what the agent makes of what her senses put to her'. Perception in contrast is comparatively unintegrated with that outlook. In a manner of speaking, then, perception remains the work of the senses.

**Acknowledgements** For comments or discussion, thanks to Carsten Hansen, Frank Hofmann, Jonathan Knowles, David Papineau, Thomas Raleigh, and Eva Schmidt; to audiences at the universities of Oslo and Luxembourg; and to anonymous referees. Special thanks to Nicholas Allott for discussions on linguistics, and to Kristoffer Sundberg and Sebastian Watzl for comments and discussions on all matters thought and sense.

**Funding** Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). The early stages of the work on the paper were supported by Research Council of Norway Grant 240645.

## Declarations

**Conflict of interest** The author has no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adger, D. (2018). The autonomy of syntax. In N. Hornstein, H. Lasnik, P. Patel-Grosz, & C. Yang (Eds.), *Syntactic structures after 60 years. The impact of the chomskyan revolution in linguistics* (pp. 153–176). De Gruyter Mouton.
- Allott, N. (2019). Scientific tractability and relevance theory. In K. Scott, R. Carston, & B. Clark (Eds.), *Relevance: Pragmatics and interpretation* (pp. 29–41). Cambridge University Press.
- Allott, N. (2020). Metacognition and inferential accounts of communication. In A. Nes & T. Chan (Eds.), *Inference and consciousness*. Routledge.
- Bartlett, F. C. (1947). The measurement of human skill. *British Medical Journal*, *1*(4510), 877–880.
- Beck, J. (2018). Marking the perception-cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy*, *96*(2), 319–334.
- Berger, C. C., & Ehrsson, H. H. (2013). Mental imagery changes multisensory perception. *Current Biology*, *23*(14), 1367–1372.
- Bird, A., & Tobin, E. (2023). Natural kinds. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Stanford University.
- Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society b: Biological Sciences*, *373*(1755), 20170341.
- Block, N. (2023). *The border between seeing and thinking*. Oxford University Press.
- Borg, E. (2004). *Minimal semantics*. Oxford University Press.
- Brandom, R. (1994). *Making it explicit*. Harvard University Press.
- Brewer, B. (1999). *Perception and reason*. Oxford University Press.
- Broadbent, D. E. (1997). Frederic Charles Bartlett, 1886–1969. *Biographical Memoirs of Fellows of the Royal Society*, *16*, 1–13.
- Burge, T. (2009). Five theses on *De Re* States and attitudes. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 246–324). Oxford University Press.
- Burge, T. (2010). *The origins of objectivity*. Oxford University Press.
- Burge, T. (2014). Reply to block: Adaptation and the upper border of perception. *Philosophy and Phenomenological Research*, *89*(3), 573–583.
- Burnston, D. C., & Cohen, J. (2015). Perceptual integration, modularity, and cognitive penetration. In A. Raftopoulos & J. Zeimbekis (Eds.), *Cognitive penetrability of perception*. Oxford University Press.
- Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, *78*(2), 275–311.
- Campbell, J. (2002). *Reference and consciousness*. Oxford University Press.
- Carruthers, P. (2006). The case for massively modular models of mind. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 3–21). Blackwell.
- Chomsky, N. (1980). *Rules and representations*. Columbia University Press.
- Cinque, G. (2013). Cognition, universal grammar, and typological generalizations. *Lingua*, *130*, 50–65.
- Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, *72*(20), 741–765. <https://doi.org/10.2307/2024640>
- Deroy, O., & Auvray, M. (2012). Reading the world through the skin and ears: A new perspective on sensory substitution. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2012.00457>
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434.
- Dwyer, S., & Pietroski, P. M. (1996). Believing in language. *Philosophy of Science*, *63*(3), 338–373.
- Evans, G. (1982). *Varieties of reference*. Oxford University Press.
- Field, S. (2004). *Career opportunities in radio*. Ferguson.
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*. <https://doi.org/10.1017/S0140525X15000965>
- Fish, W. (2010). *Philosophy of perception. A contemporary introduction*. London: Routledge.
- Fitts, P.M. (1947). *Psychological research on equipment design*. (Army Air Forces Aviation Psychology Program Research Reports) U.S. Government Printing Office.

- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Wadsworth.
- Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 840–853.
- Fodor, J. (1983). *Modularity of mind*. MIT Press.
- Frankish, K. (2018). Inner speech and outer thought. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: New voices*. OUP.
- Gao, T., & Scholl, B. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 669–684.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Allen & Unwin.
- Green, E. J. (2020). The perception-cognition border: A case for architectural division. *Philosophical Review*, 129, 323–393.
- Guarniero, G. (1974). Experience of tactile vision. *Perception*, 3(1), 101–104.
- Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39(1), 3–36.
- Hershberger, W. (1970). Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *Journal of Comparative and Physiological Psychology*, 73, 407–411.
- Higginbotham, J. (1987). The autonomy of syntax and semantics. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 119–131). MIT Press.
- Hobbes, T. (1651/1996). *Leviathan*. Cambridge University Press.
- Jensen, R. S. (1986). Editorial. *Ergonomics*, 29, 1297–1299.
- Kanizsa, G., & Gerbino, W. (1982). Amodal completion: Seeing or thinking? In J. Beck (Ed.), *Organization and representation in perception*. Erlbaum.
- Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, 18(1), 379–397.
- Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, 22(3/4), 333–347.
- Knowles, J. (2000). Knowledge of grammar as a propositional attitude. *Philosophical Psychology*, 13(3), 325–353.
- Kriegel, U. (2019). The perception/cognition divide: One more time, with feeling. In C. Limbeck-Lilienau & F. Stadler (Eds.), *The philosophy of perception and observation*. De Gruyter.
- Laeng, B., Bloem, I. M., D'Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, 131(2), 263–283.
- Lerman, H. (2012). Demonstrative content and the experience of properties. *Dialectica*, 66(4), 489–515.
- Levine, J. (2010). Demonstrative thought. *Mind & Language*, 25(2), 169–195.
- Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, 1(1), 333–353.
- Locke, J. (1698/1979). *An essay concerning human understanding*. Oxford University Press.
- Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 97(2), 267–283.
- Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. Freeman.
- Martin, J.-R., & Le Corre, F. (2015). Sensory substitution is substitution. *Mind and Language*, 30(2), 209–233.
- Mesquita, B. (2022). *Between us. How cultures create emotions*. Norton.
- Mingon, McArthur, & Sutton, J. (2021). Why robots can't haka: Skilled performance and embodied knowledge in the Māori haka. *Synthese*, 199(1–2), 4337–4365.
- Millikan, R. G. (2004). *Varieties of meaning*. MIT Press.
- Mole, C. (2015). Attention and cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives*. Oxford University Press.
- Nanay, B. (2010). A modal theory of function. *Journal of Philosophy*, 107(8), 412–431.
- Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, 172(7), 1723–1736.
- Nanay, B. (2017). Sensory substitution and multimodal mental imagery. *Perception*, 46(9), 1014.
- Nes, A., Sundberg, K., & Watzl, S. (2023). The perception/cognition distinction. *Inquiry*, 66(2), 165–195.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.
- Pautz, A. (2021). *Perception*. Routledge.
- Pence, D. E. (2021). On the capacity for vision through sensory substitution. *The British Journal for the Philosophy of Science*, 72(1), 79–103.



- Phillips, B. (2019). The shifting border between perception and cognition. *Noûs*, 53(2), 316–346.
- Phillips, B. (2021). Seeing seeing. *Philosophy and Phenomenological Research*, 102(1), 24–43.
- Phillips, I. (2016). Consciousness and criterion: On block's case for unconscious seeing. *Philosophy and Phenomenological Research*, 93(2), 419–451.
- Plato. (2014). *Theaetetus* (John McDowell & Lesley Brown, Trans.). Oxford University Press.
- Prinz, J. J. (2006). Is emotion a form of perception? *Canadian Journal of Philosophy*, 36(sup1), 137–160.
- Quilly-Dunn, J. (2020). Concepts and predication from perception to cognition. *Philosophical Issues*, 30, 273–292.
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
- Recanati, F. (2007). *Perceptual thought*. Oxford University Press.
- Recanati, F. (2012). *Mental files*. Oxford University Press.
- Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., & De Volder, A. G. (2005). The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception*, 34(7), 857–867.
- Rey, G. (2020). *Representation of language*. Oxford University Press.
- Rock, I. (1983). *The logic of perception*. MIT Press.
- Rock, I. (1997). *Indirect perception*. MIT Press.
- Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology*, 52, 453–470.
- Samuels, R. (2006). Is the human mind massively modular? In S. Rod (Ed.), *Contemporary debates in cognitive science*. Blackwell.
- Samuels, R. (2012). Massive modularity. In E. Margolis, R. Samuels, & S. Stich (Eds.), *Oxford handbook of philosophy of cognitive science*. Oxford University Press.
- Schmidt, R. A., & Lee, T. D. (2019). *Motor learning and performance. From principles to applications* (6th ed.). Human Kinetics.
- Scholl, B. J. (2005). Innateness and (Bayesian) visual perception: Reconciling nativism and development. In P. Carruthers & S. Laurence (Eds.), *Stephen Stich The innate mind: Structure and contents* (pp. 34–52). Oxford University Press.
- Searle, J. R. (1983). *Intentionality*. Cambridge University Press.
- Siegel, A., & Sapru, H. N. (2006). *Essential neuroscience*. Kluwer.
- Siegle, J. H., & Warren, W. H. (2010). Distal attribution and distance perception in sensory substitution. *Perception*, 39(2), 208–223.
- Smith, N. V., Tsimpli, I.-M., & Ouhalla, J. (1993). Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant. *Lingua*, 91(4), 279–347.
- Stanley, J. (2005). Semantic knowledge and practical knowledge. II: *Hornsby on the phenomenology of speech Aristotelian society supplementary volume*, 79(1), 131–145. <https://doi.org/10.1111/j.0309-7013.2005.00129.x>
- Stegmann, U. (2014). Causal control and genetic causation. *Noûs*, 48(3), 450–465.
- Stokes, D. (2018). Attention and the cognitive penetrability of perception. *Australasian Journal of Philosophy*, 96(2), 303–318.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description III: Grammatical categories and the lexicon*. Cambridge University Press.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.
- William, L. B., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, 6, 345–375.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.