A Naturalistic Theory of Perceptual Representation

ADAM R. LEES

The University of Sydney

I propose a naturalistic theory of representation for characterising perceptual events and their effects on language and other higher functions, intended *inter alia* to inform the metasemantic constraints on accounts of the semantics of natural languages. While it is orthodox in the cognitive sciences to characterise perceptual events as carrying information about their external causes, I argue that they should instead be characterised as carrying information determined by their role in the production of behaviour, and that this interpretation provides a basis for their semantic evaluability.

The proposed theory seeks a general characterisation of the perceptual events posited and studied by the cognitive and neuro-sciences—neurophysiological phenomena to which perceptual roles have been ascribed mainly by inferences from observed behaviour (e.g. self-reports *qua* heterophenomenology,\(^1\) or measurements of task performance). While I suspect that the theory would play some important role in analyses of the folk concepts or ordinary language associated with perception, qualia, etc., such considerations will not be the main focus here. Instead, the theory will be recommended on roughly empirical rather than roughly conceptual grounds (i.e. justified with respect to the features of the relevant processes *qua* scientific subject, rather than the semantics of ‘perception’ *qua* scientific subject).\(^2\)

In §1, I outline desiderata that the theory should satisfy given its motivations, then in §2 I prepare and present a theory of representation meeting these desiderata. I begin (in §2.1) by arguing that the best general characterisation of perceptual events for explaining their causal role in the system has them tokening states according to their behavioural contribution, where this is cashed out as *ceteris paribus* effects on the motor control system, or alternatively as disjunctions of behavioural commands that, together with motivational inputs, causally determine behavioural outputs in the absence of executive functions and learning. In §2.2, I argue that the

---

1 ‘Heterophenomenology’ is Dennett’s coinage for the practice of using a subject’s reports of her own phenomenological experiences as third-person data (1991: 72–83). This practice is ubiquitous in perceptual research (or it is, at least, a more naturalistically defensible version of what is ubiquitous in the research).

2 The theory cannot sidestep conceptual requirements entirely. In order for it to qualify as a theory of representation, it must satisfy the conceptual mandates put forward in the second desideratum D2 (§1.2).
representational contents of tokened perceptual states are specified by the identities of the perceptual states as they are characterised in §2.1, but only when these tokens affect the executive systems. Much of the support for this theory of representation comes from the view that the basic function of the executive systems is to intercept perceptual inputs to the motor control system and to facilitate learning by simulating behaviours and their perceptual feedback. Intuitively, the theory says that perceived objects and properties consist in patterns of, or constructions out of, the behavioural options made available by perceptual events to the executive systems. The theory thereby countenances strictly de dicto representations (in a sense to be explained below), and it follows—due to a background commitment to naturalism—that successful perceptual representation depends upon the matching of the features of a behavioural model representation to a natural kind. I give some reasons for doubting that such perceptual representations are ever successful in this way, and thus for doubting that an attitude of realism towards manifest types of objects and properties is justified. In §3, I briefly discuss some of the issues raised by the theory.

1 Desiderata

This section establishes six desiderata that the theory of perceptual representation must satisfy.

1.1 Naturalistic Representation (D1)

The theory of perceptual representation must be naturalistic. Since representation and intensionality have traditionally been problematic for a naturalistic-scientific worldview, I therefore need to be explicit about the kind of naturalism to which I intend the theory to conform.

I am sympathetic to the Moorean view that ordinary claims, ceteris paribus, ought to be preferred over extraordinary claims, but I do not think the burden of proof should lie ultimately with those who are sceptical of the ordinary. Rather, I think that what Fine calls “post-Moorean modesty”\(^3\) is best served by a naturalistic perspective that shifts the burden of proof finally to the anti-scientific. For the naturalist, every truth must pay its way in the currency of scientific explanation, since inter alia the scientific method is our technology for getting at the true and the real in spite of the illusion and bias inherent in many of our pre-theoretic attitudes. Thus

\(^3\) Fine 2002: 5.
wherever the manifest world of Moorean facts is incompatible with the world as posited by our best science, so much the worse for the incompatible Moorean facts.

We have positioned science as the unique arbiter of which claims are most likely to be true, and the notion of truth invoked here is ultimate and nonrelative—that is, truth *qua* correspondence with a reality that is independent of the generation of the truth bearers. The practice of science involves the (fallible) justification of statements that are prima facie about what exists in a robustly theory-independent sense, and we tend to view our best current science as the basis for all currently justified such statements. Our best current science is thus our privileged discourse of what is real in the strongest available sense of ‘real’, and any claim about reality holds water only insofar as it is a current scientific claim or bears some special relation to such claims. It is this kind of *epistemological naturalism* to which I intend the theory to conform:

\[ [EN] \]
Where \( S_t \) is the set of all and only the descriptions justified by our best scientific practices at time \( t \), and where \( T_t \) is the set of descriptions properly related to those in \( S_t \), the union of \( S_t \) and \( T_t \) is the set of claims we are justified in asserting or believing to be (independently) true at \( t \). The only types of entities about which an attitude of realism is justified at \( t \) are natural kinds, where \( \Phi \) is a natural kind iff \( \Phi \) is a type *defined by* descriptions in \( S_t \cup T_t \).

In order to get an intuitive grip on the notion of a *type*, as employed by [EN], I shall use an extensional working definition in terms of possible worlds:

\[ [\text{TYPES}_\text{e}] \]
A type \( \Phi \) is a collection of all and only those subsets of the domains of logically possible worlds \( w \) at times \( t \) that contain all and only the elements satisfying a particular description \( p \).

---

4 See Sellars on the manifest/scientific distinction (1963: 1–40).
5 Quine, ironically, does not seem to acknowledge this as a corollary of naturalism (see 1960: 3, 21–2).
6 I take this correspondence notion of truth to be the ordinary and intuitive notion (even if this consists merely in its characterising ordinary statements about truth).
7 Modulo concerns about whether or how there could ever be a unique set of such claims. It is of no great consequence if the conjunctions in \( S_t \) are riddled with disjunctions of rival claims tied for first place.
8 I should stress that these need only be *epistemologically* natural kinds, given [EN]. The term ‘natural kind’ is not meant to imply metaphysical realism (see below).
9 Cf. Lewis on properties (1986: 50–69). Unlike Lewis, I tend to think that *worlds* in the present context are best treated as logically possible states of the *actual* world, with type definitions ideally figuring in scientific predictions of the actual tokening of some types (i.e. predictions of natural states of the actual world) from the actual tokening of other types.
Given the above definition, we shall say that the description \( p \) defines the type \( \Phi \) (this is the sense of ‘defined’ in [EN]). If \( x \) satisfies \( p \) at some \( w \) and \( t \), then we shall say that \( x \) tokens \( \Phi \) (at \( w, t \)). I shall use the shorthand ‘\( \Phi s \)’ as a collective noun denoting all possible tokens of the type \( \Phi \) (at all times), and ‘\( \exists x \Phi x \)’ to mean that there is some actually occurrent token \( x \) of \( \Phi \). We shall say that \( \Phi \) is a type of \( \Gamma \) just if all tokens of \( \Phi \) also token the type \( \Gamma \). So, for example, \( \text{the electron} \) is a type of subatomic particle because it is a set of possible collections of all things meeting the definitive description of \( \text{the electron} \) (i.e. every description in which the term ‘electron’ occurs in \( S \)), and all of these things are also subatomic particles.

I am remaining neutral about the identities of the proper relations to \( S \), mentioned in [EN], however I take it that if a claim \( p \) explains or simplifies the explanation of some phenomenon by relating or providing alternative definitions for only natural kinds, then this is sufficient for \( p \) being an element of \( T \), and thus defining a natural kind itself. Accordingly, there would be tokens of the type defined by \( p \) at worlds at which \( S \cup T \) is satisfied (i.e. at nomologically possible worlds, i.e. at all worlds compatible with our best account of the actual world at \( \bar{t} \)). Therefore, if the theory of perceptual representation consists in claims like \( p \), then it is naturalistic in accordance with [EN].

It is the view of Price and others that notions like representation are not appropriately naturalistic.\(^{10}\) Price argues that what he calls “subject naturalism”—the scientific study of \( us \), the language-using, theory-having subject—is the starting position from which we would have to justify any claim about the representation of “metaphysical” reality,\(^{11}\) even the representation of metaphysical reality by science itself (this representational view of science being close to what Price calls “object naturalism”). Moreover, he argues that the justification of this kind of representationalism from subject naturalism is at least unnecessary, and perhaps even impossible. By contrast, I argue that the naturalistic study of the human subject does in fact justify the claim that perceptual states represent metaphysical (i.e. mind-independent) reality.

It is important to note that representing mind-independent reality does not imply ever successfully doing so (e.g. where representations of mind-independent conditions are natural kinds, but not the represented conditions themselves). Furthermore, the naturalistic justification of a claim \( p \) that perceptual states represent does not depend upon \( p \) itself representing representation in the

\(^{10}\) Price 2004.

\(^{11}\) I.e. theory-independent reality (Putnam 1980).
very way it describes (a self-stultifying circularity discussed by Price with respect to “object naturalism”). Given [EN], the justification of \( p \) may consist merely in showing that \( p \) is properly related to the claims justified by our best current science. Whether science itself represents metaphysical reality (i.e. whether “object naturalism” is justified) is a separate question that need not be answered here; and an answer to this question would more plausibly encounter the circularity that the claim \( p \), being about perception and not about science, avoids. Again, the kind of representationalism to which Price and others are opposed is one that explicitly commits to semantic relations \textit{really holding} between theoretical claims and aspects of metaphysical reality independent of those claims. I have only insisted that claims properly related to our best current science deserve to be \textit{treated} as metaphysically real. My neutral stance here can be seen as a methodological form of metaphysical quietism, but one that at least treats the ontological claims inherent in our scientific practices as the privileged discourse of the metaphysically real. Except for this methodological character, and for the fact that I am not committing to expressivism, this makes my approach similar to Simon Blackburn’s quasi-realism about ethics (but as applied to scientific discourse only), and to a methodological version of Putnam’s empirical realism.

Consequently, my goal here is only to argue—in line with \( D1 \)—for a theory of representation that is compatible with the epistemologically naturalist, metaphysically quietist methodology as just stated. This does not require arguing in support of the metaphysical realist, but does require contradicting the blanket antirepresentationalist. If you are a metaphysical realist, and you agree with [EN], then interpret me as arguing that you ought to believe or assert that perceptual representation is metaphysically real for \textit{more} than just pragmatic reasons. If you are a staunch theory-relativist or antirepresentationalist, and you agree with [EN], then simply interpret me as arguing that the theory of perceptual representation is compatible with our

---

12 Price 2004. This charge of circularity is not unrelated to Putnam’s (1980) “just more theory” response to constraints on reference, and is probably an instance of the scepticism-favouring “disputational deadlock” discussed by Lewis (1999).

13 Blackburn 1993. Macarthur and Price (2007) advocate a \textit{global} application of Blackburn’s quasi-realism. This resembles my methodology in just the respect that it applies to scientific theories, but differs in that this is just a \textit{special} case of its global application, the quasi-realism is not purely methodological, and the authors commit to global non-cognitivism. The [EN] position is that scientific discourse is the \textit{privileged} discourse of metaphysical reality (e.g. [EN] would advocate antirealism with respect to ethical discourse if it were not properly related to science). Put differently: no claim even gets to be \textit{quasi-true} at \( t \) unless it relates to \( S \), in the right way. There is then the separate (and perhaps unanswerable) metaphysical question about whether the quasi-true claims are also genuinely true. Nevertheless, we are able to \textit{justify} realism or antirealism with respect to a discourse because its being related to \( S \), in the right way is a necessary condition for its posits being justifiably \textit{treated} as metaphysically real.

14 Putnam 1980.
scientific discourse, and thus consists in valid moves in the “metaphysical realism” language game. Arguing for either of these interpretations means arguing for a conclusion about the representational status of our best current science.

1.2 Basis for Evaluability (D2)

\[ \text{[STATES]} \] A state is a type of event in (i.e. a type of activity in, or a type of spatiotemporal configuration of the parts of) a system (i.e. a mechanism, or an organised collection of interacting parts).

In line with the above definition and those in D1, the theory of perceptual representation must give an explanatory definition \( p \) of states of the perceptual system (a neurophysiological system with certain behavioural outputs, as discussed in §1.4), such that these states are natural kinds, and such that any event satisfying \( p \) (i.e. any token of a perceptual state) satisfies the criteria for representation. These criteria for representation are as follows. Firstly, if any state \( \Psi \) is representational, then it must somehow define a type \( \Phi \) such that any tokening of \( \Psi \) represents that \( \Phi \) is tokened. This requirement is neutral as to whether we account for either or both of (i) the de re representation of some actual, indicated individual as being a certain way (such that the representation defining \( \Phi \) is partly indexical, or is otherwise guaranteed to be satisfied by a unique individual at the same \( w \) and \( t \) at which \( \Psi \) is tokened) or (ii) the de dicto representation that there is some actual individual that is a certain way (such that the definition of \( \Phi \) is not indexical, etc.). A de re representation, whether successful or not, implies that there is necessarily some actual \( x \) that it is representing as being a certain way, whereas a de dicto representation represents both that there is some \( x \) and that it is a certain way, such that if it is unsuccessful then there is not necessarily any such \( x \).

Secondly, \( \Psi \)s must be capable of misrepresentation: there must be a \( w \) and \( t \) such that the state \( \Psi \) is tokened at \( w, t \) and yet the type \( \Phi \) that it defines is not tokened at \( w, t \). This capacity for misrepresentation ensures that the relevant states are semantically evaluable, or assessable for accuracy (or inaccuracy) with respect to the obtaining of \( \exists x \Phi x \) for some type \( \Phi \).\(^{15}\) Accordingly, a

\(^{15}\) The disjunction problem for the unqualified causal theory of content arises because it fails to deal adequately with misrepresentation (see Fodor 1992).
tokened representational state $\psi$ is veridical or successful just if $\exists x \Phi x$, where $\psi$s represent that $\exists x \Phi x$.\textsuperscript{16}

Determining what $\psi$s specifically represent—for example, the identity of the type $\Phi$ where $\psi$s represent that $\exists x \Phi x$—will depend, I think, on what $\psi$s have in common by virtue of which they have the same content. Instead of locating a natural kind of representation relation that the tokens of different representational states bear to different represented types of external conditions, we want to locate a natural property of perceptual events that justifies characterising them as representations in the first place—call these properties the ‘evaluability’ or ‘$\mathcal{E}$’-properties of the perceptual events-cum-tokens. The $\mathcal{E}$-properties are those properties in virtue of which different types of perceptual events are evaluable with respect to different sets of possible conditions, and thereby token different representational states. $D2$ requires that tokens of the same representational state $\psi$ have the same $\mathcal{E}$-property, which is their common basis for evaluability given which they mutually token $\psi$.

1.3 \textbf{AVOIDING THE CONTENT IDENTIFICATION PROBLEM ($D3$)}

I want the theory of representation to steer clear of certain assumptions that are commonly made by both philosophers of content and cognitive scientists; it must avoid a sequence such as the following:

1. A theory of the content of $\psi$s is proposed, according to which $\psi$s represent that $\exists x \Phi x$.
2. The theory relies on a definition of the type $\Phi$ that in turn depends on the tokening of $\psi$.

The theory proposed in step 1 has failed. It has effectively told us that what $\psi$s represent is (at least in part) simply what $\psi$s represent. This is an instance of what I am calling ‘The Content Identification Problem’ (CIP). CIP is usually encountered when the represented type $\Phi$ is a perceptually represented type of external condition, identified by intersubjective agreement in observation. For this reason, it is related to the problem of simply assuming that manifest (or Moorean) types are also natural kinds. However, CIP compounds this problem—in CIP, the assumption that ordinary types are real in this way is relied upon in a theoretical account while at the same time begging a central question that the account is meant to answer.

\textsuperscript{16} Note that $\exists x \Phi x$ covers the contents of both de re and de dicto representations defining $\Phi$, since we are characterising the de re nature of a representation as arising from its defining $\Phi$ differently (i.e. indexically, or like a definite description whose satisfaction is necessitated by the tokening of the de re representation).
A philosopher would encounter CIP if she were to advocate a causal theory of perceptual content-determination, and then straight away go on to speak of perceptual representations of apples, chairs, blueness, etc., all the while taking it for granted that these manifest objects and properties are available to be the causes of tokened representational states. Without further argument, we would have no basis for supposing that apples, chairs, blueness, and so on, are among the natural kinds that could qualify as causes. The philosopher has relied upon her own perceptual representations of such entities as evidence that they are representation-independent causes, and this is simply question-begging. If her goal was to explain how we come to represent such things as apples and chairs, rather than to merely characterise how perceptual contents (whatever they might be) are determined, then she has failed to achieve this goal, for the reasons outlined above. If she did in fact seek this latter goal, then it would simply be an open question whether any given manifest, represented object or property was a natural kind (i.e. a nomologically possible kind) that might serve as the cause of its own representation.

Cognitive neuroscientists are similarly faced with CIP. Suppose that a scientist conducts an experiment in which various objects are placed in subjects’ fields of vision, and their brain activity is recorded. The data show a significant correlation between brain state $Ξ$ and visual fixation on “cubes” (irrespective of size, angle of presentation, etc.), and the scientist takes this as evidence supporting the conclusion that a tokened state $Ξ$ encodes, is responsive to, or is caused by cubes “in the world”. If the scientist is right, then her identification of the covariant stimulus as a cube depends upon her own tokening of $Ξ$. So if the scientist sought to determine the relationship between brain states and types of external stimuli then she has failed to do so—her evidence supports only the conclusion that $Ξ$ is tokened when people agree that, or act as if, they are seeing “cubes”.

CIP tends not to be seen as posing a problem, because it is assumed that the relevant tokened representation is at least sometimes successful in its representing and therefore represents some natural kind $Φ$, and moreover that mere agreement in the obtaining of an instance of $Φ$ is sufficient evidence for its obtaining. However, it should be clear that this sort of assumption undermines any proposed explanation of how manifest types such as cubes come to be represented. We shall not make such an assumption.\textsuperscript{17} If our theory says that $Ψ$s represent that

\textsuperscript{17} Unfortunately, this means having to steer clear of misleading but otherwise intuitive examples of representation, and mostly persevering with the use of Greek letters standing for the representing and represented types.
∃xΦx, then it must point to an account of the definition of Φ that is independent of Ψ. In particular, given [EN], it must point to an account of the definition of Φ that is properly related to science.

1.4 Application (D4)

The theory of perceptual representation must be applicable to human-instantiated systems, as sketched in figure 1 (overleaf). It is critical to the goal of explaining the impact of perception on language that the theory be applicable to systems with human-like executive functions—systems responsible for behaviours associated with reasoning, language, long-term planning, self-monitoring, etc. The causal-functional arrangement in fig. 1 has been simplified as much as possible given present purposes, and it is beyond the scope of this paper to argue convincingly for it. I do however think that it is representative of findings in the cognitive and neuro-sciences.

Interpreting the sort of research on which fig. 1 is based—and interpreting fig. 1 itself—requires being mindful of CIP (in addition to neutralising the theory of content often implicit in the research). For example, the object system (mutatis mutandis for the property systems) is not defined with respect to an unequivocal kind of function or content, but rather with respect to a kind of evidence—convergent from multiple disciplines within psychology and neuroscience—that speaks to the functional specificity and dissociability of the processes responsible for certain behaviours (and these processes could be interpreted in one or another specific functional or intentional way).\(^{18}\) The evidence suggests that some sorts of functionally specific processes subserve the successful performance of tasks requiring the explicit identification of objects qua objects, and of the “object-making” properties that are interdependent with these object identifications—for example, tasks that involve object identity during sensory contact (necessary for object tracking), object identity between disconnected sensory episodes (necessary for re-identification and object permanence), and identity of form, shape, and so on. But the evidence depends upon experimental stimuli only identified as being objects (and having certain object-making properties) by the experimenter. As covered in §1.3, the ubiquitous inference from this sort of evidence to the conclusion that the relevant processes in the brain function to identify or recover features of these external objects from ambiguous sensory data runs afoul of CIP.

\(^{18}\) E.g. see Kanwisher 2010.
In order to avoid the problem, we must maintain the default position on the underlying object system—and, mutatis mutandis, on any of the property systems—that it has only been shown to be necessary for successfully demonstrating agreement with the object identification performance of the experimenter, or at best with the norm of a sample or population. Unless I mention otherwise, then, I will use ‘object identification’ and ‘property identification’ to denote successful agreement with the relevant norm, without begging the question about the independent scientific status of the types of external objects or properties so identified. Likewise, the successful “recognition” of an identified object or property as being the member of a learnt category will also pertain, by default, to successful agreement with a norm.

Figure 1. A simplified model of functional divisions and causal relationships within the human neuro-functional system.
With this in mind, some features of the general causal-functional architecture presented in fig. 1 should be explained, particularly with respect to the perceptual systems. I shall be using ‘memory’ and ‘learning’ in the restricted sense of individual learning—in other words, learning that is not part of universal early development relatively unaffected by environmental factors that are developmentally irrelevant—and I will use ‘innate’ in the correspondingly broad sense. As shown in fig. 1, activation of the detectors (sensory receptors) is the cause of events in what I have called the ‘property systems’ and the ‘object system’, which are instantiated in widely distributed regions of the cortex. The property systems are necessary for abstract identifications of colour, motion, pitch, odour, etc., that are independent of the learnt discrimination and generalisation of these abstract properties into categories (e.g. into the category “red”). While some property systems may receive inputs from multiple sensory modalities, most property systems are specific to the input of a single modality (e.g. colour properties and the visual modality). By contrast, the object system is at least partly modality-nonspecific, and is necessary for mediating learning-independent identifications of distinguishable and cohered abstract object forms, on the basis of rich input from the property systems (primarily of the visual and tactile-haptic modalities). Object identifications (and identifications of object-making properties) are necessarily made on the basis of at least some prior property identifications, although there appears to be no unique set of sufficient property identifications for any given abstract object identification. So while object identifications would be impaired with each loss of a property system, they would survive at least some loss in richness of input. Without an intact object system—and hence the ability to identify objects abstractly—it would still be possible to make property identifications, as evidenced in the visual case by sufferers of visual form agnosia. Farah describes these agnosics as evidencing “stuff vision in the absence of thing vision”, such that the

19 E.g. the systems subserving each sensory modality are located roughly in the visual cortex, auditory cortex, etc., and extend to the respective sensory association cortices. Object system behaviours have been associated mainly with, e.g., ventral occiptotemporal regions and the lateral occiptal complex (see Grill-Spector & Malach 2004).

20 Note the distinction between being able to respond differentially to colours in the absence of learnt colour categories like “red” (some of which will happen to fall within the normative range of “red”), and being able to respond differentially to reds and non-reds on the basis of this category. The acategorical colour identifications are strictly independent of individual learning, but there is evidence that once categories are learnt they can facilitate disambiguation and identification of both properties and objects at the learning-independent perceptual systems level (e.g. see Bar 2003).


22 Spelke 1993; Farah 2004.
intact property systems “yield a kind of rich but formless visual goo”.\textsuperscript{23} \textit{Inter alia}, it follows that we can perceive or identify uninstantiated properties, or properties that we do not identify as holding of any object.\textsuperscript{24}

Lastly, many of an individual’s 	extit{learnt} categories of perceived objects and properties are formed in the process of learning her native language. However, the capacity to make the acategorical, perception-level object and property identifications on which this learning is based is \textit{innate}. Some of this capacity appears to be present from birth,\textsuperscript{25} while the rest appears to follow a near-universal developmental schedule that is already complete by the time a child begins to acquire her native language.\textsuperscript{26} Without the perceptual systems, the normal learning of the familiar object and property categories (e.g. as might be associated with words like ‘chair’, ‘wolf’, ‘red’) would be impossible, and with it the recognition of members of these categories by their perceived characteristics.\textsuperscript{27} Moreover, a competent adult speaker who came to lose one of her perceptual systems would be unable to recognise a learnt category member on the \textit{basis} of the identification that was mediated by that lost system (e.g. without the basic ability to identify colours, one would not be able to recognise instances of the property category “red”). If the learning of an object or property category involved various, distributed perceptual identifications, then the ability to recognise a member of this category would of course survive some level of impairment of the property systems. However, the complete loss of the \textit{object} system would make \textit{object} category recognition impossible—this kind of visual form agnosic would have to resort to circumlocutive guesswork based on the perceived confluence of properties that had \textit{already} been associated with the object category in question, via learning.\textsuperscript{28}

The above considerations inform some of the functional architecture modelled in \textit{fig. 1}, and especially the causal priority and dissociability among the property, object, and memory systems.

\textsuperscript{23} Farah: 18–19.

\textsuperscript{24} This is one reason why identifications of “object-making” properties, which are not dissociable from object identifications and tend to be modality-nonspecific, are functionally separated into the object system.

\textsuperscript{25} Streri & Spelke; de Haan & Nelson 1999; Meltzoff & Borton; Valenza et al. 1996.

\textsuperscript{26} Spelke.

\textsuperscript{27} E.g. see Soja et al. 1993; Spelke.

\textsuperscript{28} This is the going theory for explaining \textit{visual form agnosic} recognition performance data (Farah).
1.5 Selection by Behaviour (D5)

Evolutionary teleosemantic accounts latch the semantic evaluability required by D2 onto the natural biological functions of brain states, so that the semantic success or failure of a tokened state just is the success or failure of the token at fulfilling the function for which the state was selected. However, teleosemanticists differ over how to characterise the sort of function for which a state would have been selected—that is, they differ over the identity of the $E$-properties—and hence over what counts as the success or truth conditions for such a state.  

Teleosemanticists who are sympathetic to the interpretations of content in the cognitive sciences tend to view the selectional function of a cognitive state as giving a kind of semantically evaluable causal-informational content (which I will call ‘teleoindicative’ content). Accordingly, a state $\Psi$ of the relevant kind is viewed as having been selected so that its tokens will be caused by (or covary with, track, or indicate) instances of some external natural kind $\Phi$.  

$\Psi$s are taken to de re represent that their causes $x$ are of the kind $\Phi$ that $\Psi$ was selected to track, such that they successfully represent their causative conditions $x$ only if $x$ is in fact an instance of $\Phi$, and de re misrepresent $x$ as being of the kind $\Phi$ only if $x$ is not of the kind that $\Psi$ was selected to track.  

Compared with the alternative teleosemantic account discussed below, the teleoindicative approach is somewhat more congenial to a computational model of cognitive processes given which cognitive processes are algorithms operating upon the tokens of semantically atomistic states. The computational model is, in turn, commonly employed by cognitive scientists.

Despite this, we might think it better to view representational states as having been selected for their contribution to the production of behaviour under natural conditions, rather than for the purpose of tracking such conditions. As MacDonald and Papineau put it:

…biological functions are in the first instance a matter of effects…. [A mental representation’s] function will be to enable the consumer mechanism to achieve its end by gearing behaviour to circumstances.

---

29 MacDonald and Papineau (2006) distinguish the following rival kinds of teleosemantics as input-based and output-based, respectively.

30 This is the sort of function that undergirds the causal-informational teleosemantics recommended by Neander (e.g. 1995).

31 Dretske (1991) calls a system of such states a “natural system of representation”.

32 The classic example from cognitive science (and, aptly, from perception) is Marr’s computational theory of vision (1982).

33 MacDonald & Papineau: 4.
Nervous systems basically consist in links between detectors (e.g. sensory receptors) and effectors (e.g. motor neurons). The maximally general evolutionary function of every part of every nervous system (qua nervous system) is to figure ultimately in that system’s production of fitness-enhancing behaviour in response to environmental causes (i.e. to figure in the fitness-enhancing activation of effectors in response to some type of activation of the detectors), not to merely indicate these causes in isolation (i.e. in the absence of effector activation). If nervous systems could not have generated such a fitness advantage in terms of behavioral interaction with the environment, then they would have provided no traction for selection, and could not have evolved. To make this difference in priority clear, consider that if the tokens of a state $\Psi$ would be able to reliably produce a specific behaviour via effectors in all and only $\Phi$ conditions without receiving any afferent indication of $\Phi$ conditions from detectors, then $\Psi$ could still be selected for if doing this were fitness-enhancing. However, if $\Psi$s would be able to reliably indicate or track $\Phi$ conditions, but would not be able to influence the behaviour of the organism on this basis, then $\Psi$ simply could not evolve no matter how fitness-enhancing it would have been to track $\Phi$ conditions if only $\Psi$s could have influenced behaviour. So far as natural selection is concerned, then, every state $\Psi$ at every level of every nervous system tracks the causes of its tokens only instrumentally, in the process of contributing to the behaviour that is fitness-enhancing in the presence of these causes.

For this reason, inter alia, some teleosemanticists view a representational state $\Psi$ as having been selected for the function of contributing to a specific type of behaviour under a natural kind $\Phi$ of condition. Accordingly, $\Psi$s de re represent that the condition $\chi$ in which they are producing behaviour is of a kind $\Phi$, where $\Psi$ was selected to be tokened and to make the relevant behavioural contribution given $\Phi$-conditions. Thus, $\Psi$s de re misrepresent $\chi$ as being $\Phi$ just if $\chi$ is not of this selectionally-relevant kind. I will call this the ‘teleoeffective’ view.\(^\text{34}\)

A popular criticism of teleological accounts of content is that they are not very intuitive and (most importantly for our purposes) not very explanatorily useful. The notion that the species-wide, brain-mediated states under investigation (i.e. perceptual states) have been shaped and constrained by natural selection is uncontroversial; but the idea that the kind of content had by these states is identical with evolutionary functionality seems to do little to serve the cognitive

\(^{34}\) Millikan’s (1984) proper functions are teleoeffective functions, as are those undergirding Papineau’s (1993) teleosemantics.
scientist or the conceptual analyst, and would furthermore prevent us from meeting D6 (outlined below). However, the constraints imposed on the perceptual systems by natural selection are important to consider when we are looking to characterise perceptual events. The discussed insight behind the teleoeffective view is a sound one—these evolutionary constraints will primarily have to do with selection on the basis of behavioural effects, at every level of the system. Even if a representational characterisation does not appeal directly to the selectional functions of perceptual states, any characterisation of the perceptual systems will have to be compatible with their having been selected for their behavioural contributions.

1.6 Narrowness and Accessibility (D6)

Teleosemantic theories seem to provide no place for narrow content in human thought and language—that is, contents that are determined by or that supervene on only the intrinsic properties of the system as compared with the extrinsic properties that it has by virtue of its context (such as the having of a certain selectional history). On unqualified teleological accounts, if we have a system S with representational states (as determined by S's selectional history) at a time t, and also its “Swampman” system S* at t (i.e. a system that is a perfect intrinsic match of the first system S but has appeared spontaneously in a puff of smoke at t) then S* fails to have any representational states at t because it has no selectional history, even though it matches S in every natural intrinsic detail investigable at t.36 As such, unqualified teleological contents are purely broad, and have no narrow components. The -properties they offer as the basis for the evaluability of states are relational properties, namely selectional functions.

Whether or not this result is counter-intuitive does not concern us here. What does concern us is that a motivating goal of our theory is to offer synchronic causal explanations of the effects of perception on language and the executive systems. Fodor has argued roughly that since such causes must be local, if a naturalistic theory posits contents that occupy meaningful causal-explanatory roles then those contents must supervene on intrinsic properties, and not historical or contextual properties of the system.37 I agree with Fodor. In fact, I think we need something even stronger:

35 E.g. see Braddon-Mitchell & Jackson (1997), and Pietroski (1992) for worrying implications of the teleoeffective approach.
36 The notion of a Swampman system is based on Davidson's thought experiment in his 1997.
Accessibility

If the theory determines that $\Psi$ represents that $\exists x \Phi x$ for some specific $\Phi$, then the content of a token of $\Psi$ is accessible to the system, where this implies that:

- The token has certain specific effects on the executive systems

  -and-

- The token’s having these specific effects on the executive systems is autonomously, synchronically explained by its representing that $\exists x \Phi x$ for a specific $\Phi$ (i.e. by its tokening of the state $\Psi$).

Note that a theory could meet Fodor’s narrowness condition, and yet merely consist in using representational categories to name local physical or neurophysiological kinds. Such a project would not be putting the notion of content to autonomous theoretical use, and would invite conceptual confusion in the context of scientific inquiry. Since we are not looking for the referents of folk concepts, a mere identity thesis of this kind would seem to serve little theoretical purpose. Instead, Accessibility requires that the tokens of states with different contents differentially affect the executive systems—particularly language—and that differences between the tokens of these states in their effects on the executive systems are best synchronically explained by the very fact that the states they token differ in what they represent. Intuitively, we want an account according to which perceptual events can be described as carrying “information” about their contents, and it is the use to which this information is put by the executive systems that best explains how these systems are affected by those perceptual events. In order for the carrying of this information to provide such an explanation, it must still be cashed out somewhere among the intrinsic, causal properties of the system, and it follows from this that the perceptual contents must be narrow (i.e. supervene on these intrinsic informational properties). Accessibility also requires that contents explain autonomously—explanations by content cannot simply reduce to explanations by local causal properties that, together with physical laws, determine the “causal flow” from the perceptual events to the executive systems. Therefore, contents must be identified by properties of the system that are intrinsic, but also nonlocal or organisational.

Importantly, Accessibility does not imply that a system is able to generate accurate reports of what its tokened perceptual states represent. In other words, Accessibility does not require “conceptual” contents per se (but see note 56).
Accordingly, facts about how a system came to have a particular state $\Psi$ (facts that are difficult to determine and may even be themselves indeterminate) merely constitute a contingent back-story for the content of $\Psi$s that is attributable only from outside the system—they do not fix the identity of that content. Knowing what the tokens of a perceptual state $\Psi$ represent will allow us generally to predict the effects of $\Psi$s on reasoning and language, because it will imply that $\Psi$s convey unique information to these systems.

The failure of many functionalist theories of content to fill these explanatory, Accessibility-type roles is an explicit reason given by Fodor for avoiding such theories, and of sticking with his Asymmetrical Dependency theory of causal-indicative content. By contrast, Dretske seeks to extend his teleoindicative account of representation so that causal-indicator functions might yield Accessibility-satisfying contents and thereby provide the kinds of ultimate reasons for the actions of agents that we take to characterise beliefs and desires.

On Dretske’s extended account, a “reason-giving” state $\Psi$ represents its external causes as being of the type of cause that was responsible for $\Psi$s acquiring causal involvement in behaviour via learning. This sounds a bit like Accessibility: if the acquired, local causal paths from the reason-giving $\Psi$s to the motor control system would have ended up the same, yet the $\Psi$s happened to have had different types of external causes and thus different shared causal-informational properties, then the same reinforcement contingencies would have failed to hold and $\Psi$s would not have been assigned the same content by learning. Thus, the content of $\Psi$s seems to provide an autonomous explanation for the behaviour they cause. As Dennett notes, however, there is no substantive difference between diachronic content-determining functions (qua $E$–properties) being assigned by natural selection, and synchronic content-determining functions (qua $E$–properties) being assigned by developmental selection. In both cases, the functional properties that determine the contents are historical-relational, and so the contents in both cases are purely broad. It follows that these contents also fail to be accessible to the system,
and so cannot figure in the right kind of generalised explanation of the causal roles of the tokened representations. Note that in the above scenario, learning merely provides a “back-story” explanation for both the content and the relevant causal role of the state. Even though this story insists that you cannot have one without the other, the causal role does not directly depend on the content; rather, they are interdependent.

One way to see this is to note that a Swampman system $S^\star$, which intrinsically matches a system $S$ with acquired Dretskean reason-giving representations, has no reason-giving representations of its own because it lacks a learning history. Nevertheless, any intrinsic counterpart $x^\star$ in $S^\star$ of some putatively reason-giving representation $x$ in $S$ has all the same effects (on counterpart systems) as $x$ has in $S$. The causal influence of any acquired representation post-learning in $S$ is the same as its counterpart in $S^\star$ even though $S^\star$ underwent no learning. The downstream systems themselves “cannot tell” (do not respond dependently upon) whether they are in $S$ or $S^\star$, thus cannot tell whether or not a perceptual event tokens a state that has been linked to behaviour via learning, and thus cannot tell whether such an event has reason-giving content or not. Their responses to intrinsically matched events will themselves be intrinsically matched, irrespective of learning history, and so learning does not add any information that is accessible in $S$ but not in $S^\star$. Developmental teleoindicative contents, then, fail to be accessible.

Other advocates of teleological theories might be already willing to accept the non-accessibility of contents just by consideration of the intrinsic causal properties of the tokening system, quite apart from any commitment to identifying contents with selectional functions. Dennett—one stripe of teleosemanticist—seems to think that a view satisfying something like Accessibility veers too close to the idea of a homunculus reviewing mental contents in some Cartesian theatre in the brain. He alludes to its implausibility—of contents autonomously explaining causal roles—by invoking the brain’s intrinsic properties:

---

43 In fact, it is even worse than this. Not just teleoindicative content, but any kind of causal-indicative content fails to satisfy Accessibility (see discussion of the quotation below and also §2.1).

44 ‘Cartesian theatre’ is Dennett’s (1991) derisive coinage for a “command headquarters” where “everything comes together” for consciousness in the brain (compare the role I will end up giving to the executive systems in perceptual representation).
It seems plausible enough that discrimination of stimuli and doing something about them are perfectly separable. But [...] in the brain, discrimination of afferents according to their significance just is the production of efferent effects in differential response to afferents.45

By going on to accept a teleosemantic story, then, Dennett does not appear to have much to lose. But while the input-output characterisation of neural processes in the second line of the quotation is fair, note that the first line assumes that the contents qua “significance” of afferents—the contents that would have to be accessible to the system—would have to be information about the (functional) external causes of the afferents. As Dennett points out, there does not seem to be anywhere to locate this information about causative external stimuli in the system itself, such that the system might function in a way that is better synchronically explained by the use of this information than simply by a causal-physical description (even if this description were amenable to a diachronic explanation invoking natural selection). Of course, this is not the only possible way of defining narrow representational contents. The theory of perceptual representation proposed below is perfectly compatible with Dennett’s characterisation of the intrinsic nature of neural processes, while at the same time satisfying Accessibility in accordance with our motivating goals.

2 The “Behavioural Model” Theory of Perceptual Representation

In building up to the proposal (in §2.2) of a theory of perceptual representation meeting the desiderata just outlined, I shall begin by recommending a definition of perceptual states in terms of the behavioural contributions of perceptual events.

2.1 The “Behavioural Advocacy” Characterisation of Perceptual Events

The first step is to note that there will be no general type-type correspondence between natural kinds of causes and natural kinds of perceptual events. Dissociable patterns of variance in the activity of the perceptual systems will not just magically get to line up with nature’s joints. Since natural selection built these systems, the extent to which perceptual activity will causally track natural kinds will be roughly the extent to which this was fitness-enhancing for ancestral systems. In line with D5, then, this is the extent to which it was instrumental in making behavioural contributions that paid for it in ancestral inclusive fitness. Natural selection is also a notorious miser, designing systems on a shoestring budget. At some point in the evolution of an

45 Dennett 2010: 82–3 (my emphasis).
adaptation, diminishing returns will tip any possibility of further improvement from being a fitness-benefit to a fitness-cost, and the selection pressure driving its “research and development” will relax. Therefore we should not expect perceptual systems to have a level of causal-indicative discrimination much in excess of sufficiency for subserving adaptive behaviours.

Given all this, we should be sceptical about natural kinds of perceptual events and their systemic roles mapping onto natural kinds (i.e. onto types of external conditions that are already defined in the relevant scientific theory of those conditions). Our only guarantee is that perceptual activity will indicatively differentiate between natural kinds of external conditions given which it was persistently fitness-enhancing to behave differentially. (Inevitably, these collections of kinds will then be shrouded in the causal-indicative noise of freeloader kinds to which selection was blind.) Inter alia, this means that we cannot characterise natural kinds of perceptual events cause-first, in general—e.g. as by starting with an existing scientifically-defined type Φ of external condition, and then defining the perceptual state by causal covariance with Φ. The types of external conditions defined by causal covariance with perceptual patterns will be among any number of conflated and amended hodge-podges of natural kinds—that is, the gerrymandered types which it was adaptive for behaviour to discriminate between but not within, together with their respective scatterings of freeloaders. The science positing the underlying kinds (e.g. physics) will itself offer no principle for predicting the perceptually relevant hodge-podges from the perceptually irrelevant; for that, we must look to the system itself. So the teleoeffective theorists are right about one thing: the only principled and general way to characterise the causative external conditions of perceptual events is to first suppose that the perceptual events are distinguishable by natural kinds selected for their contribution to specific behaviour under particular circumstances.46 We can then say that the tokening of such a “teleoeffectively” defined state Ψ will be caused by any condition of the hodge-podge of natural kinds given which Ψ was selected to be tokened together with the related freeloader kinds. Defined as a unitary type in its own right, this hodge-podge of causes of Ψs will not necessarily be a pre-existing natural kind, defined by an existing scientific theory; rather it will be a type that is

46 This is the only general characterisation, because for the reasons given we cannot just inductively generalise from cases of covariance to the next closest external natural kind; so our only alternative would be to exhaustively catalogue each causative external condition we observe (and this is clearly neither general nor predictive).
biologico-physically defined relative to the teleoeffective function of $\Psi$—the type of cause of $\Psi$s will be defined by our new theory of perception.

These biologico-physical types of external causes, as predicted by teleoeffective functions, would still only be of contingent relevance to the operation of the system. Hold fixed the behavioural contribution for which $\Psi$ was selected, while varying the historical circumstances given which it was selected to make this contribution, and you change the hodge-podge of natural kinds whose instances cause it to be tokened, without changing any intrinsic property of the system that we might wish to synchronically explain (or at least no intrinsic properties that are causally downstream from $\Psi$s). According to the method we used to define the perceptual states, we would now have to say that by changing the inputs that lead to $\Psi'$s behavioural contribution, we have made it so that perceptual events can no longer be classified under the state $\Psi$, and we would have to replace $\Psi$ with some other state $\Upsilon$ defined by the new input-output configuration. But the causal role occupied within the system by the relevant perceptual events would not have changed with the change in their classification from $\Psi$ to $\Upsilon$. The tokens of $\Psi$ and $\Upsilon$ tell the system the same thing—they convey nothing about their different external hodge-podges of causes. All that matters to the system itself is the causal role (i.e. the behavioural contribution half of the teleoeffective ledger), and this is shared by the perceptual events in both cases. The contributions to behaviour made by perceptual events will be entailed by the intrinsic causal architecture of the system. Varying the behavioural contribution made by a perceptual event—unlike varying just its external causes—necessarily means altering the direct causal influence of that event on the system, and many intrinsic details of the system along the way. For this reason, I suggest that the best general characterisation of perceptual states that can be employed in synchronic explanations is a characterisation by type of causal role in the production of behaviour, irrespective of the circumstances under which this role tends to be occupied, or given which the role was selected to be occupied. This property of having specific effects on the system forms a natural kind of perceptual event that is necessitated by intrinsic properties of the system, and can be described as information that is carried by a perceptual event within the system.

\[47\] I.e. if we defined the content of $\Psi$s relative to any kind of causative condition, then this content would be inaccessible. As the quote from Dennett in §2.6 touched upon, the causing of $\Psi$s by actual causative conditions past or present cannot synchronically explain $\Psi$s' internal effects.
It is possible to distinguish further the kind of contribution made by a perceptual event—we will end up describing them as carrying disjunctions of motor instructions (or behavioural options) for downstream use in the system. Recall that the perceptual processes we care about are innate or at least biologically prepared (§1.4). These processes are the basic building blocks of the object and property categories we come to learn. The capacity for basic perceptual identifications is necessary for the formation of learnt perceptual categories, and is causally prior to the perceptual recognition of learnt category members. Therefore, the behavioural contribution (or information carried) by a tokened perceptual state will generalise across those individual systems that share a selectional history (i.e. that share their early architecture). This will be the broadly innate information that comes to be used in some way by memory and the executive systems in order to respond to the changing circumstances of the individual. So the behavioural contributions of perceptual events will not include the less direct contributions to behaviour that come to be made via their effects upon memory and the executive systems, and that vary between individuals on the basis of individual histories.

It follows that the kind of behavioural contribution with which we should define the perceptual states will be compatible with a system lacking executive functions and any individual learning (i.e. a naive system of a kind that could be thought of as ancestral to human systems). The immanent behaviour of such a system would in principle be predictable, irrespective of the system’s history of inputs, from (i) the current pattern of detector activation or a complete description of activity in the perceptual systems, and (ii) the current (innate or biologically-prepared) motivational state. I have generalised all motor control functions to a single system—MCS—which takes inputs from both the motivational system and from the perceptual systems and produces behaviour by causing patterns of effector activations (fig. 1). Thus, the contribution to behaviour that is specific to the perceptual systems will be recoverable from the total MCS activity (in the absence of input from memory and the executive systems) by subtracting the contribution made by the motivational system.

48 Perceptual states are occupying a belief-like role in something like a Humean theory of action, since, in the absence of learning and the capacity for behavioural inhibition (a recently evolved, executive capacity) perceptual events will directly cause behaviour in conjunction with the motivational input. However, since the discussed perceptual and motivational states are basic and innate, and there are no executive systems for reasoning, long-term planning, and behavioural modulation, this would not be an instance of a theory of action according to which the action would be readily classifiable as agentive.
The motivational states might also be defined with respect to their causal contributions to MCS, and for similar reasons as in the perceptual case. However, the contributions of their tokens just to the activity in MCS must be far simpler than the contributions made by perceptual events. Perception plus motivation (in the absence of learning and inference) must produce behaviour by which the organism can navigate the world (in only those ways that were reliable across generations) to obtain currently required, biologically-salient goals. The motivational component here is expressed entirely by ‘currently required, biologically salient goal’. Suppose that the total motivational contribution is caused by the tokened state HUNGRY, the protracted tokening of which happens to be reliably caused by internal detector activations that in turn happen to be reliably caused by nutritional deficiencies. Selection has designed the system so that when a perceptual state \(\Psi\) is tokened that was a reliable indicator of tokens of an ancestrally-relevant kind of food, MCS will use the total information from the perceptual systems (not just from the token of \(\Psi\)) and the information from the token of HUNGRY, to initiate behaviour via the effectors (behaviour of a kind that, from outside the system, as it happens, will probably move the organism closer to the food source that has probably caused \(\Psi\) to be tokened). We could describe the way HUNGRY functions with a rule like ‘if \(\Psi\) is tokened, then tell MCS to execute \(\Psi\)-oriented approach behaviour’, and this might even be a good theoretical move to make. But at the moment we are only interested in characterising the HUNGRY token’s causal influence on MCS, so that we might subtract it out to get the contribution being made by the perceptual input. At best, then, the information the token of HUNGRY contributes to MCS—namely, ‘execute \(\Psi\)-oriented approach behaviour’—is a simple command that is parasitic upon the behavioural contribution made by the tokened perceptual state \(\Psi\), and the “approach behaviour” effector programmes subsequently activated by MCS. Since the specific “approach” behaviour that will be relevant on a given occasion will depend on the rest of the perceptual input to MCS, the effect of a tokened motivational state like HUNGRY can only be to cause MCS to execute effector programmes of a certain very general type (e.g. the “approach” type), which can then be applied relative to any perceptual input.

By contrast, perceptual states have a lot of work to do. We can distinguish perceptual events roughly by how they are used by MCS in the production of behaviour, or, in other words, between (i) those that are directly relevant to motivational contributions to MCS and that anchor the ongoing selection of behavioural programmes with respect to that motivational input.
(intuitively, events caused by goals or threats relative to which behaviour must be organised; just the \( \Psi \) token, in the current example); and \((ii)\) the rest—namely all of the other contributing perceptual events, responsible for determining which types of behavioural programmes would be generally appropriate given the relevant perceptual anchor and the motivational input (in this case the HUNGRY token’s ‘execute \( \Psi \)-oriented approach behaviour’ command). Intuitively, this latter kind of perceptual input will tend to gear behaviour to the distance and orientation of the food source probably causing the anchoring perceptual input, and to intervening obstacles, etc.\(^{49}\)

Perceptual events would need to be consistently usable by MCS no matter what the current motivational input. If HUNGRY were not currently tokened, then the causal contribution of a \( \Psi \) token to MCS might need to be able to occupy a role in behavioural selection instead of the anchor role of behaviour-orienting (externally-speaking, perhaps the food token likely causing the \( \Psi \) token is, in some other motivational context, an obstacle that must be moved out of the way in pursuit of some other biologically-salient goal). The same state \( \Psi \) is tokened either way, because perceptual activity is unaffected by the concurrently tokened motivational state on its way to MCS, and thus in both instances the contribution made to behaviour via causal influence on MCS (the contribution by which perceptual states are differentiated) must be the same. Therefore, tokened perceptual states must be carriers of all the information that could be required by either role, and the distinction between these roles must result from which part of the information is used by MCS and perhaps also how it is used in light of the motivational input.

We can further clarify this general behavioural advocacy of perceptual events by again adopting a perspective external to the system. Imagine that the individual’s path to the food token (the probable cause of the \( \Psi \) token) is obstructed in some way, so that the system would have to negotiate these obstacles to successfully approach the food token. It would have been fitness-enhancing, as it happens, for perceptual states to be (probabilistically) causally covariant with whatever gerrymanderings of kinds of external features would have been reliably useful for just this purpose, namely for behavioural manipulation in the pursuit of currently required, biologically-salient goals. As discussed, the highest level of causal-indicative richness or fineness of discrimination among these external features would then correspond with the level of fineness of discrimination among these external features would then correspond with the level of fineness.

\(^{49}\) We could tell the same kind of story with, e.g., a motivational state FEAR and a perceptual type \( \Psi \) whose tokens are reliably caused by predators, such that a token of FEAR would contribute something more like ‘execute \( \Psi \)-oriented avoidant behaviour’ to MCS, and then MCS would use the \( \Psi \) token’s contribution for the purpose of anchoring “avoiding” effector programmes (execute: run, execute: hide, etc.).
of behavioural interaction with the environment that was ancestrally adaptive (and no higher, in general). Internally, what this would amount to is a corresponding richness of behavioural contributions by perceptual states (the intrinsic half of the teleoeffective ledger). Since perceptual states are defined by these behavioural contributions, there will be as many definable perceptual states tokened as there are tokenings of causative hodge-podges that are behaviourally indivisible—that is, as many definable perceptual states as there are gerrymanderings of natural kinds between which it was not fitness-enhancing to behaviourally differentiate.

Consider a perceptual state \( \Psi \) whose tokens happen to be externally caused by conditions of a single behaviourally indivisible hodge-podge of natural kinds, such that \( \Psi \) makes what we shall call an ‘atomic’ behavioural contribution to the activity in MCS. \( \Psi \)'s would need to tell MCS what types of effector programmes it would be appropriate to execute given just some \( \Psi \) token as input, irrespective of the motivational input—that is, as it happens, every type of behavioural option whose benefit to fitness was \textit{ceteris paribus} constant over every reliable type of ancestral cause of \( \Psi \). So \( \Psi \)'s will advocate these types of effector programmes, and MCS will execute a programme of one of these types should it be “motivated” to engage with \( \Psi \). Intuitively, \( \Psi \)'s will point MCS to a disjunction of types of effector programmes. Externally-speaking, conditional upon motivational input requiring MCS to make use of the contribution by \( \Psi \), this would tend to amount to orienting types of behavioural patterns with respect to a certain external location relative to the body, as well as disposing behaviour towards local engagements appropriate to the expected environment\(^{50}\) (i.e. appropriate to environmental conditions of the type gerrymandered by the conferring of equal fitness value to the very behavioural engagements advocated by \( \Psi \)).

We can therefore characterise the causal influence on MCS of a tokened atomic perceptual state as a disjunction of commands to execute effector programmes of certain types. Externally-speaking, and more intuitively, the behavioural contribution made by the token of a perceptual state is a list of general options for producing appropriate behavioural engagements with the ancestral environment (i.e. \( \Psi \)’s tell MCS “do something like \textit{this or this or this or...}”).

Since perceptual states are defined simply by this behavioural contribution, tokened atomic perceptual states that simultaneously contribute to MCS can be thought of as jointly tokening a \textit{composite} state that makes a single causal contribution to MCS—that is, a disjunction

\(^{50}\) Internally, this would just mean having programme types that happen to be appropriate to \textit{egocentric location and probable type of cause} as options advocated by \( \Psi \).
of the component atomic disjunctions of behavioural commands, leaving all of the atomic behavioural options on the table for use by MCS in light of the motivational input. Furthermore, a tokened atomic state—or more often a tokened composite of atomic states—will also routinely be the cause of the tokening of a higher-level atomic perceptual state that makes its own distinct behavioural contribution. As activity progresses towards MCS from the detectors of the system, this will happen progressively more often. This is what happens when tokened states in the property systems cause tokenings of states in the object system. In such a case, tokens of a composite property state $\Psi$ will cause the tokens of an atomic object state $\Xi$ to contribute their defining disjunction of behavioural commands to MCS, while $\Psi$s simultaneously have their own distinct effects on MCS. The “blasphemous”, Marrian way of describing this kind of occurrence (at least in the visual case) would be to say that information about three-dimensional objects is recovered by the tokened higher-level atomic state from information about colour or motion, etc., which is carried by the tokened lower-level composite. As we have seen, however, this sort of indicative information is not carried downstream within the system, and also risks confronting CIP. We have instead characterised the conveyed information in terms of related sets of behavioural options.

Atomic and composite perceptual states, then, are both uniquely defined as follows:

\[ PT \]

\begin{itemize}
  \item For any event in the perceptual systems $x$ and perceptual state $\Psi$, there is some set $M_\Psi = \{ \Lambda_1, ..., \Lambda_n \}$ of types of effector programmes (i.e. effector activation patterns) such that $x$ tokens $\Psi$ iff, were $x$ to affect MCS, then $x$ would \textit{ceteris paribus} directly causally contribute to the increased likelihood of the execution of a programme of every type in $M_\Psi$, and, in the absence of input from memory and the executive systems, if MCS were to contiguously execute such a programme then $x$ will have directly causally contributed to this execution.

  \item \textit{Equivalently}: $x$ tokens $\Psi$ iff there is a unique set $M_\Psi = \{ \Lambda_1, ..., \Lambda_n \}$ of types of effector programmes such that $x$ conveys the disjunction of behavioural commands $\nu_\Psi \approx \langle \exists y \Lambda_1 y, \text{ or } ..., \text{ or } \exists z \Lambda_n z \rangle$ to MCS.
\end{itemize}

\[ PT \] defines a natural kind of perceptual event, based on intrinsic causal properties of the system, which is compatible with the proper selectional considerations (D5) and steers clear, at least so
far, of CIP (D3). The teleoeffective theorist about content would say that the perceptual states thus characterised qualify as representational states (even in the absence of memory and the executive systems). However, so far all we have is a way of grouping together causal pathways according to their common endpoints. Per Dennett, describing such a system as representing originating external causes might be useful for hermeneutics and prediction, but as yet nothing more. Locating the representation of external conditions somewhere in this causal story would be a difficult task, and accomplishing it would have little or no synchronic explanatory value.

2.2 Representation by Behavioural Model

We are now poised to develop a theory of perceptual representation on the basis of the definition [PT] of perceptual states just given.

Armed with [PT], we can get a quick sense of what happens to our “nonexecutive” system when it undergoes learning. The teleoeffective raison d'être of a memory system, I think it is safe enough to say, is to gear behaviour to the inter-individually and inter-generationally variable circumstances that cannot be tracked by natural selection—a feat that can be accomplished only to the extent that it was fitness-enhancing cross-generationally and could thereby itself be the subject of selection. We can think of learning as consisting in coordinated, widely-distributed processes that modify and augment innate functional architecture by altering the causal relationships between existing systems (one example of a result of this would be adding to or fine-tuning the motor programmes that MCS might execute upon perceptual advocacy of a behavioural type). As in fig. 1, this can be functionally rendered as a discrete memory system with reciprocal causal interactions with existing systems, altering causal pathways on the basis of (reinforcing) input from the motivational system. Once we factor in the memory system, the internal complexity of detector-effector (externally: stimulus-response) pathways will quickly ramp up, and the history of the system’s inputs will start to matter for predicting its behaviour.

It is instructive to think of the executive systems as having their evolutionary origin as a functional outgrowth of MCS and the memory system. Despite being a hallmark of humanity, and possessed of an almost endless functional fecundity, I contend that the selection of their basic functional plan was initially driven rather indirectly by improvement upon memory’s

51 Dennett 1989.
function of gearing behaviour to individual circumstances, namely by \((a)\) steepening the
instrumental learning curve, and \((b)\) decreasing the fitness cost of instrumental learning.\(^{52}\)

The gearing of behaviour by instrumental learning involves trial and error, which of
course poses some risk to the organism. The errors that do not kill you might make you stronger
in the long run, but they could in fact kill you. Even if the errors take some toll, however small,
then it would be much preferable to acquire the improvement in a less risky manner. The
synchronic functions of the executive systems permit what amounts to the virtual or simulated
execution of effector programmes—virtual trials and errors—under the inter-individually variable
circumstances ordinarily requiring the gearing of behaviour through instrumental learning alone.
This process would of course rely upon some input from learning for the purposes of prediction
and extrapolation—it would, after all, be a variety of inductive inference (e.g. we can imagine a
homunculus version of the executive systems thinking to itself “based upon past experience with
behaving similarly in similar circumstances, what might happen if in my current situation I
behaved thusly?”). Since it would not obviate the need for standard instrumental learning, this
executive process would at best decrease the number of trials required to reach the best fit of
behaviour to the circumstances, and therefore only decrease the fitness cost of instrumental
learning. Moreover, as Hume acknowledged,\(^{53}\) the output of this kind of process will be fallible,
and so will ultimately require the harsh guidance of the actual environment to which behaviour is
being geared, and not merely of the approximations of the executive systems.

This sort of functionality requires that the executive systems be well-connected—it
requires that they bear reciprocal causal connections to MCS, but also to every other system. In a
nonexecutive system, if the concurrent perceptual and motivational input to MCS at a time were
such that MCS would execute an effector programme of a given type, then the causal path from
activity at the detectors to activity at the effectors, accounting for memory states and the tokened
motivational state, would be inevitable. So the executive systems must be able to inhibit these
automatic causal pathways to behaviour by causally influencing activity in MCS, pending the

\(^{52}\) Of course, the fitness benefit of learning processes must have outweighed their fitness cost overall; a further reduction in any
fitness cost would then increase the overall fitness benefit.

\(^{53}\) Hume 2007/1748, §IV.
results of the virtual trials.\textsuperscript{54} Furthermore, this executive behavioural inhibition must be able to be caused by input to the executive systems from the perceptual and motivational systems, so that the inhibitory effect may anticipate executions by MCS.

Activity in the executive systems must also be affected by the same (or sufficiently many of the same) perceptual and motivational events as those that would have caused activity in MCS, so that possible responses to this input by MCS may be simulated. Moreover, the effector programme types that the perceptual events would have advocated to MCS must be identifiable and considerable by (i.e. must differentially affect) the executive systems, both for the purpose of trialling appropriate effector programmes, and for providing the initial cueing input for the process. While space constraints preclude a detailed account, these virtual trials would consist in something like (a) perceptual, motivational, and memory inputs to executive systems causing the inhibition of inputs to MCS, then (b) the tokening of other perceptual states and associative memory states, caused by the precipitating inputs and the types of effector programmes that are on trial, until (c) the tokening of an associated motivational state signals an expected reward, and (d) the executive systems cause the relevant type of effector activation via MCS.

This simulating function of the executive systems became progressively more central to human ancestors, so that modern humans spend much of their time with the executive “foot on the break” of the motor control system, in a near-constant mode of prediction and anticipation, setting long-term goals with respect to which behaviour is controlled and fine-tuned. The basic trialling process—a kind of behavioural planning—creates a vast potential for flexibility and for further functionality.\textsuperscript{55} But for present purposes, the most important feature of this process is that it supports a representational characterisation of the tokened perceptual states that causally influence it, as follows:

\textsuperscript{54} This is compatible with various neurocognitive theories and studies regarding executive function. The part of the executive systems involved in behavioural inhibition seems to be instantiated by prefrontal areas, and the relevant part of MCS seems to be instantiated in the basal ganglia (Chambers et al. 2008; Mostofsky & Simmonds 2008).

\textsuperscript{55} E.g. the feedback loops that permit this executive function by causing states to be tokened in the perceptual and memory systems—“autostimulations” whose ultimate purpose is to facilitate instrumental learning via trialling—could plausibly underpin the cognitive functions of relational learning and working memory (the latter based partly in the recapitulation and maintenance of activity in the perceptual systems), as well as self-monitoring, attention, imagination, and so on.
The $\mathbf{E}$-property by virtue of which perceptual events token a representational state $\Psi$ is the property of sharing specific *ceteris paribus* MCS effects (i.e. the property of carrying a specific disjunction of behavioural commands) while also having a causal influence on the executive systems.

Recall that by definition [PT], all possible tokens of a given perceptual state $\Psi$ share the property of having a specific causal influence, *ceteris paribus*, on MCS—the property of carrying the specific disjunction $\nu_\Psi$. Thus $\Psi$s affect the executive systems (i.e. are involved in virtual trials of effector programmes) just if they also token a specific *representational* state that represents that $\exists x \Phi x$ for some specific $\Phi$. In such a case, the identity of the represented type $\Phi$ is specified by $\nu_\Psi$. So a token of the state representing that $\exists x \Phi x$ just is a token of the perceptual state $\Psi$ when this token is affecting the executive systems.

As just discussed, in order for the executive systems to trial effector programmes in light of occurrently tokened perceptual states, their activity must be differentially responsive to the behavioural options that *would* have been advocated to MCS by those same perceptual tokens. Since the carrying of this behavioural information is what defines the perceptual states, the executive systems are thus *differentially responsive* to the tokens of different perceptual states. If the executive systems are responding to the tokens of such states in the first place, then these perceptual tokens qualify as tokening corresponding representational states *in virtue* of the specific behavioural information they carry, as just stipulated. Since the contents of these states are also specified by this behavioural information, it follows that the contents are accessible ($D6$).

Unlike Dretske’s reason-giving representations, downstream systems can *always* tell a representation from a nonrepresentation, because by definition the only downstream systems are the executive systems, and if they notice a perceptual event then, by definition, this event counts as a representation. Additionally, downstream systems can always tell one representational content from another, because this content is internally specified.

Tokens of the same perceptual representation have the content they do in virtue of the $\mathbf{E}$-property they share, namely the property of carrying of a specific $\nu$ while influencing the executive systems. Moreover, the fact that the tokens of *different* representational states *differentially* affect the executive systems is synchronically explained by the fact that they *differ* in their $\mathbf{E}$-properties—representational $\Psi$s and $\Xi$s both affect the executive systems by definition, yet they carry *different* disjunctions of behavioural commands (i.e. they would have affected MCS
differently), and for this reason the executive systems must be set up to respond to them differently for the purpose of trialling effector programmes. If we hold fixed the local causal paths from $\Psi$s to the executive systems, but change the effect that $\Psi$s would have had on MCS (and thus the disjunction $v_\psi$ defining the $\Psi$s), then the executive systems will be \textit{differently affected} by $\Psi$s because they will trial different effector programmes in response. The internal structure of the executive systems, which determines the activity configurations in those systems upon a contribution by $\Psi$s, will vary as we vary the effect $v_\psi$ of $\Psi$s on MCS, and this in turn will vary the content of $\Psi$s, since the identity of this content is determined by $v_\psi$.

This story invokes only the intrinsic, causal properties of the system. Since the theory of representation just given also applies when we have a system as shown in fig.1 (D4), and is based on a characterisation of perceptual states that is compatible with D3 and D5 (i.e. the naturalistic definition [PT] of perceptual states in §2.1) we will satisfy all of our desiderata as long as the determination of contents by the $\mathcal{C}$-properties (i.e. as long as D2) can be given sufficient support.

So how would merely having some effect on the executive systems, together with carrying a specific $\nu$, determine the content of a perceptual event? For the purpose of trialling effector programmes, the change in executive systems activity immediately following perceptual input—some input-exhaustive disjunction of behavioural commands $v_\Omega$—will have to be contingent upon (or “take into consideration”) \textit{all} of the information in $v_\Omega$, and thus be sensitive to any differences between different exhaustive inputs. This much follows from the differential effects of tokens of different perceptual states, but intuitively we ought to think that this is because the complete perceptual information at a time will causally determine which parts of this information will subsequently figure in behavioural planning (i.e. determine the subsequent causal pathway in the executive systems). The executive systems must initially “consider” \textit{all} of the information coming from perception at once. Imagine any given $v_\Omega$ as a rich matrix of behavioural options available to the executive systems, with each $v_\Omega$ distinguished by its unique causal profile with respect to MCS effector activation. This matrix is like the system’s representation \textit{to itself} of its own, constantly shifting behavioural dispositions. Patterns in the behavioural options in this matrix, coinciding with causal relationships holding between tokened perceptual states of the perceptual systems (e.g. between property tokens and object tokens, as discussed in §2.1) will define loci of behaviourally significant relationships—clusters formed by the advocacy of related “anchoring” programme types whose purpose, \textit{inter alia}, is to orient
behavioural engagements with respect to the body. The perceptual input, in other words, provides rich behavioural models to the executive systems that set the trialling processes into motion.\textsuperscript{56}

When the subject and the experimenter agree in their innate perceptual identification capacities—for example, when a subject performs any task that requires correctly identifying and distinguishing a “cube” qua abstract three-dimensional form—activity in the subject’s object system is tokening particular perceptual states. For simplicity, we shall pretend that this involves just a single tokened perceptual composite $\Psi$ in all cases of cube-identification. If the subject instantiated a nonexecutive system, then the tokening of $\Psi$ would make a direct and inexorable contribution to behaviour when accounting for motivational state. Changing this automatic behavioural contribution by perception in the nonexecutive system would have required a gradual process of new learning. But because the subject of the experiment has executive systems, these systems are able to inhibit the automatic contribution of $\Psi$s and to use this contribution in flexibly responding to novel perceptual input (e.g. the manipulations of the experimenter) by planning a behavioural response that is predicted to be more rewarding given the novel input. The experimenter, and the rest of the subject’s fellow Homo sapiens, share broadly similar perceptual and effector systems by virtue of sharing a selectional history. Thus whatever is causing the tokening of $\Psi$ in the subject is likely to be causing a perceptual input to the experimenter’s executive systems—a behavioural model—lawfully related to the one provided in the subject by the $\Psi$ token. The fact that the subject succeeds in flexibly engaging with and responding to the “cube”, as represented by the experimenter, is due to the subject’s executive systems trialling behaviours on the basis of behavioural dispositions from the perceptual systems that are species-specific, and thus lawfully related to the concurrent dispositions of the experimenter.

When the subject flexibly responds to, or talks about, the features of a “cube” (its surfaces, edges, position, size, motion, anticipated three-dimensionality, etc.), her behaviour (even when mediated by the learnt categories of language, since these are built from her basic identification

\textsuperscript{56}This picture resembles Tye’s (1995; 2000) account of perceptual contents as map-like and behaviour-informing, and functionally poised to influence the concept-forming machinery. Here, the maps are plots of types of behavioural programmes, and the executive systems are poised to plan and execute behaviour on their basis. However, unlike Tye, we are not immediately concerned with phenomenal experiences. And since, on our account, perception represents only when it is in fact interfacing with the executive systems, it seems less appropriate to describe perceptual representations (as contrasted with the lower-level states on which they are based) as nonconceptual (or analog rather than digital, in Dretske’s terminology (1981: 135–68)).
capacities) is caused ultimately by the effects of certain kinds of behaviour-advocating information from the tokening of the composite $Ψ$. For example, the perceived surfaces of the cube, in three dimensions, will consist in the modelling of points of potential behavioural anchoring and engagement. Relative to the total perceptual input, $v_Ψ$ models a locus $X$—a kind of cluster of related behavioural options. $v_Ψ$ “says” to the executive systems: “there is some $X$ for which an MCS programme $Λ_1, u$ would be appropriate, or $Λ_2, z$ or ...”. By virtue inter alia of the particular regions of the effector system required for anchoring types of programmes with respect to $v_Ψ$, and the relationship between the cluster of behavioural options that form $X$ on the one hand, and certain perceptual constants on the other (e.g. behavioural input originating from kinaesthetic detector activations, etc.), $X$ has a specific egocentric position in the “potential behaviour world” of loci modelled by the total perceptual input. In other words, $X$ is represented as being independent of the system representing it. The behavioural locus $X$ causally influences behaviour, heterophenomenological reports, and the general use of language in such a way that $X$ is taken to be a three-dimensional, mind-independent object out there in a world that we directly perceive, such that we are free to interact with this object (i.e. such that the executive systems are trialling behaviour with respect to $v_Ψ$). While the causal story about perceptual representations is cashed out in terms of models of potential behaviour, the perceptual representations themselves are transparent—the behavioural mode of representation is not itself represented.

So what it is for us to perceive, for example, that there is a cube over there with such and such objecty properties—at least when it comes to the causes of our deliberate behaviour and language—is for certain kinds of behavioural options to be advocated by perceptual events to the executive systems. Since by supposition $Ψ$s provide all of these options in the “cube” case (i.e. they have the right complex of effects on the executive systems), $Ψ$s represent that there is a specific $X$ (a cube) egocentrically located at some position $z$, with “objecty” properties $F, G, ...$

---

57 Somewhat similar connections have been made between the potential for behavioural engagement and the representing of objects as mind-independent, but they are almost always made in the context of conscious experiences (with which the present theory does not directly engage). E.g. such a connection is made by Husserl (2001/1901) in terms of the changes to phenomenology expected with movement, and carried through by Merleau-Ponty as the motor-intentional “readiness” to engage with objects (2002), although both of these cases invoke explicit inference in addition to the experiential dimension (both unjustified, in my view). There are also similarities to be found in Peacocke’s “positioned scenario” contents (1999). Perhaps the most relevant is Cussins’ (1990) “construction-theoretic” notion of spatial perception as the having of particular motor abilities (bar the fact that it focusses on experiences and that the mentioned abilities, qua abilities, are not semantically evaluable).
A Naturalistic Theory of Perceptual Representation

(where \(F, G, \text{etc.},\) are properties relating to motion, form, and so on, at least some of which are necessary for the representation of \(\mathbf{X},\) and which consist in potential behaviour advocated by \(v_0\)). In other words, \(\Psi\)s represent \(\exists x \Phi x,\) where \(x\) is of the type \(\Phi\) defined by the description just given, namely of there being something at the egocentric location \(z\) of a certain size and form, etc., as modelled by the behavioural options advocated by \(\Psi\)s.

Note carefully that \(\Psi\)s—and indeed all behavioural model representations (and all belief-like memory types formed entirely on the basis of such representations) only de dicto represent. This is just as well, since a de re representation requires pinning the representation of \(\exists x \Phi x\) onto a particular external individual \(x\) (which may or may not fully satisfy the representation defining \(\Phi\)). A popular way of doing this is by identifying \(x\) as the cause of \(\Psi\)’s tokening, and we have seen that such causes, just by being causes, are not directly accessible to the system as contents. Since they only de dicto represent, \(\Psi\)s will happen to represent their causes only if these causes contingently satisfy the behavioural model representation defining \(\Phi\) (irrespective of their status as causes). Since \(\Psi\)s are transparent representations of mind-independent cubes (in the present example), \(\Psi\)s misrepresent only if there in fact is nothing egocentrically located at position \(z\), with cube-properties \(F, G,\) etc. (i.e. only if \(\neg \exists x \Phi x\), irrespective of what is causing them.

According to [\(\text{EN}\)], an attitude of (mind-independent) realism is only justified with respect to natural kinds, and so the tokens of a perceptual state representing \(\exists x \Phi x\) never successfully represent if the type \(\Phi\) defined by behavioural model is not a natural kind. Given the present example, this would mean that our best current scientific descriptions do not define any types \(X_1, X_2, \ldots,\) such that \(X_1 = F, X_2 = G, \ldots,\) where \(F, G,\) etc., are the perceptually represented cube-properties. But then it is implausible, I think, to expect that perceptual representations successfully or veridically represent the world in general.

Recall that properties like \(F, G,\) and so on depend upon the behaviour-advocating disjunctions conveyed by the tokens of certain perceptual states, and the tokening of the most causally-discriminating of these states is only guaranteed to covary with hodge-podges of natural kinds among which it was not fitness-enhancing for behaviour to discriminate. To successfully represent, the behavioural options defining a perceptual state would have to be such that the behavioural model constructed out of them by the executive system—full of details, and at a level of discrimination, of relevance primarily to adaptive behavioural potential—yielded a representation that was at least uniquely related in the proper way with the definition of some
natural kind X. By my lights, since these representations are constructs out of contingently useful behavioural distinctions, the possibility of such a contingency holding, in general, is a remote one. If such contingencies fail to hold then, given the present example, we would be justified in saying that there strictly are no cubes (since cubes would be nomologically impossible) and that there are only cube representations together with the hodge-podge of natural kinds that cause them.58 Perhaps this should not surprise us, since we already have an inkling that objects of a manifest, perception-based type seem to be treated by common sense in a schematic or Platonic manner that is not easily reconciled with the verdicts of physics, and that such objects tolerate or conflate differences among the physical causes of their representation in a way that gives rise to the paradoxes of vagueness such as the problem of the heap.59

The situation for the properties is similar. I also want to say that (non-objecty) property identifications are mediated by perceptual representations that $\exists x \Phi x$ for some $\Phi$, via the construction of models for behavioural planning by the executive systems. Properties such as colour—to take the well-worn case—make sense primarily with respect to behavioural organisation; that is, primarily with respect to distinguishing among different targets of potential behaviour. A popular story of the evolution of trichromatic vision in primates talks about the selection pressure to distinguish ripe from unripe fruit, and fruit from foliage, such that the rich panoply of human colour identifications has come along for the evolutionary ride as by-products of this parochial origin.60 As C. L. Hardin has pointed out, there appear to be no natural kinds of intrinsic properties that map to identically-coloured things, and yet we ascribe colour properties to things as if they were properties intrinsic to objects.61 On the behavioural model theory, this is because discriminations between colours are just discriminations between specific kinds of anchoring behavioural commands that will, as it happens, conflate those natural distinctions among their causes that were irrelevant to the fitness-benefits of behaviour. According to the evolutionary just-so story, all colour identifications will have stemmed from the fitness-enhancing contribution of these property types to contrastive distinctions for the purpose of anchoring effector programmes—that is, they will have stemmed from the kinds of anchoring

58 An ordinary case of someone hallucinating a cube, then, would best be described as that person representing a cube under conditions that would not normally cause such a representation, or that are not of the right biologico-physical type of cause.
59 Eventually, I hope to apply these sorts of considerations to such paradoxes.
60 E.g. Osorio & Vorovyev 1996; Regan et al. 2001.
commands that help to form and differentiate behavioural loci when modelled by the executive systems. Externally-speaking, these anchoring commands would have involved the differential orienting of behaviour with respect to ripe vs. unripe fruit, and fruit vs. foliage, etc.

We should also expect these behaviour-anchoring contrasts to conflate distinct natural kinds that were irrelevant to their fitness-value, such as different types of surface reflectance profiles given which the same behavioural contrasts was rewarded in the same way by the nutritional value of fruit ripeness etc., or differences in ambient lighting that, again, made no difference to the fitness-value of the relevant behavioural contrasts (i.e. to the rewarding of the contrastive behaviour by the nutritional benefit of ripe fruit). Also, we should expect that which colour identifications are made on a specific occasion will depend, not on any specific differences between external hodge-podges of causes, but on whichever behaviour-informing contrasts were the more fitness-enhancing given the present profile of tokened “colour” perceptual states. The well-studied visual phenomena of colour constancy and the context-relativity of colour perceptions seem to bear out these sorts of expectations.62 On the present story, all other nuances to human colour identifications, including the colour-based illusions, are simply by-products of the relevant behavioural contrasts.

As the symptoms of visual form agnosia suggest, property identifications that would ordinarily help to cause object representations (and would thereby be associated with identified objects) will still be made in the absence of object representations (see §1.4). We appear to be capable of identifying disembodied redness, motion, tones, textures,63 etc., both by task performance and report, as located at various positions relative to the body. Likewise, reports of visual after-effects and illusory edges, etc., will seemingly place these items “out there in the world”, even if this is resistant to the formation of a more robust kind of credence (e.g. even if the executive systems do not trial behaviour effectively on their basis). I suggest that in the absence of the contribution of a relevant object representation to the behavioural model—a situation probably too rare to be noticed by selection—a colour-representing state $\Psi$, tokened in the visual property system, will form a sort of quasi-locus of behavioural relevance in virtue of advocating an egocentrically-located anchor-point at which it is itself the only salient information.

---

62 E.g. see Hardin.

63 As it happens, visual form agnosia do seem to be able to identity objects by touch (Farah).
In such cases, $\Psi$s represent that there is $\times$ (redness) at egocentric location $\xi$ of general size $H$, etc.\textsuperscript{64}

Under normal circumstances, the contribution by $\Psi$s will cluster with a behavioural locus representing an object such as a cube, in which case its anchoring function will contribute to a composite representation that there is an $\times$ (a cube) at location $\xi$ with objecty properties $F$, $G$, ..., and the property redness. By virtue of the defining contribution of $\Psi$s being to isolate and contrast distinct behavioural loci, redness is thereby represented as an intrinsic property of the loci which it helps to construct,\textsuperscript{65} at the respective egocentric positions of those loci, and thus as an intrinsic property of the transparently represented, mind-independent objects that it helps to define.\textsuperscript{66}

Following Hardin, it does not seem appropriate to say that the ascription of an intrinsic colour property to a thing on this basis—resting on an ancestrally-relevant behavioural distinction—maps onto natural kinds of external features, especially given evidence that there are no such kinds. And since ordinary colour ascriptions rest on the bedrock of perceptual representations, we should take the position, as naturalists holding to [EN], that there strictly are no colours\textsuperscript{67} (independently of whether there are in fact any objects like cubes that could be coloured). Instead, there are only colour representations together with the hodge-podge kinds that cause them. The behavioural model representations that subserve the other property identifications also threaten those properties with a similar fate.

3 Science and Semantics

Casting the perceived Moorean world into doubt is not the end of the world. It does not mean that we cannot augment perception by reinforcing memory states—with the aid of our executive systems, including the language system—capable of forming the basis for less dubious

\textsuperscript{64} On the “behavioural contrast” story of colour perception, however, even the disembodied representation of redness will require other concurrent colour representations.

\textsuperscript{65} It is also represented as a universal, since any $\Psi$ token affecting the executive systems will represent an instance of redness.

\textsuperscript{66} Why does Jackson’s (1982) Mary not experience redness? While we are officially steering clear of qualia, it is tempting to think that it is because Mary’s perceptual systems are not sending the right composites of contrasting behavioural options to the executive systems, even though Mary’s memory states enable her to make the relevant identifications. Suppose that there is nothing else to colour experiences beyond our theory’s functional account of colour representation. Then if Mary already knows everything there is to know about this process, and thereby knows the success conditions of the “redness” representations she is unable to have, then it seems Mary lacks a kind of ability rather than any extra knowledge (cf. Lewis 1999).

\textsuperscript{67} Recall that [EN] only mandates a negative position on the rational acceptability of colours, in light of the implausibility of a proper relation between colour representations and scientific descriptions.
ontological claims.\textsuperscript{68} The best method of accruing such memory states, as enshrined in [EN], would be through reinforcement via the practice of testing our observations according to the scientific method. Nevertheless, the behavioural model theory does require special policing of the distinction between the manifest and the scientific, and interpretations of scientific results tend to be enthusiastically careless in this regard.

We had established that the Content Identification Problem would have to be avoided, and this required agnosticism about whether manifest types were in fact scientifically defined independently of their perceptual representation—in other words, whether perceptual events represented external covariants of a natural kind. The present theory of perceptual representation gives us positive reasons for doubt, because it seems prima facie implausible that identities will hold between natural kinds and the modelled types of represented conditions. It might be objected that represented types will necessarily \textit{supervene} on natural kinds—that, necessarily, changes in tokened perceptual representations will imply changes in the external conditions causally impinging on the perceptual systems, and that since these conditions will belong to some natural kind or other, the representations thereby “properly relate” to the scientific descriptions defining the natural kinds. To see why this is not enough on its own, consider that the behavioural model theory is compatible with the existence of a \textit{wronghead}.

Suppose that Rachel’s cube and sphere representations define natural kinds—the kinds \textit{cubes} and \textit{spheres}, respectively. Suppose also, for the sake of simplicity, that the scientific descriptions defining these kinds preclude the co-occurrence of cubes and spheres. We now mess with Rachel’s internal wiring, so that the detector activations that used to cause perceptual events that send cube-modelling behavioural commands to the executive systems, now cause perceptual events that send sphere-modelling behavioural commands, and vice versa. This makes Rachel a \textit{wronghead}—now her cube representations covary with \textit{spheres}, and her sphere representations covary with \textit{cubes}.

Both kinds of representation define natural kinds by supposition, but we have now made it so that instances of these natural kinds \textit{would never cause their own representation} in Rachel. At least that is what we must say on the behavioural model theory. But notice that Rachel’s cube and sphere representations do \textit{supervene} on their new natural causes. A change in the truth value of

\textsuperscript{68} We tend to call these memory states ‘beliefs’—representations of a different kind than those formed immediately from perceptual inputs.
‘Rachel represents that there is a cube nearby’ reliably implies a change in the truth value of ‘there is a sphere nearby’. Yet surely we should not conclude from this that Rachel’s cube representations are successful. Rather, given the behavioural model theory, we should say that Rachel’s cube representations are systematically misrepresenting spheres as cubes. In order to make this distinction—and I think the behavioural model theory forces us to do so—we must acknowledge that facts about the representing events themselves (i.e. their carried disjunctions of behavioural commands) constrain what kinds of relations to science will count. This is the result of contents being both accessible and represented de dicto. Because contents are wholly determined by organisational properties of the system and do not reduce to local causal influences on the executive systems, external covariants with those causal influences (even if they were of a defined natural kind) are not enough to determine the identities of the contents. As a result, the supervenience of representations on those causes is not enough to properly relate them to science and to justify the reality of what they represent.

It follows that we ought to be cautious when speaking of the reduction of a manifest to a scientific type, or the tracking of a natural kind by perception, or the recovery of natural features by perception. As with the cases of cubes and colours, we may find it best to talk separately of the relevant representations and the hodge-podges of natural kinds that cause them. This caution would need to be exercised in paradigm cases such as the relating of the sensation of heat and molecular agitation, colours and reflectance profiles, object forms and cohered collections of molecules, and folk descriptions (or surface characteristics) and referents. Caution with respect to this last case pertains to how we ought to approach a semantics for natural languages, to which the behavioural model theory of perceptual representation was intended to be of some help. If we hold to the theory, then we will need to revisit the relationship between (for example) the referent of ‘water’ and the narrow content or primary intension associated with the recognition of water by inhabitants of Earth and Twin Earth. Such narrow contents or primary intensions would have to be closely related to the behavioural model representations of these

---

69 This does not imply that the behavioural model theory makes something like inverted spectra possible. Recall the suggestion that colour identifications be defined by contrastive distinctions among loci of potential behavioural engagements. A “colour” version of the present example would retain these contrastive relationships. The theory does, however, give us the freedom to add unique behavioural relevance to distinct colour identifications (i.e. content that further individuates colour representations) and thereby, inter alia, to honour folk intuitions about the possibility of inverted spectra.

70 Putnam 1979.
inhabitants. The behavioural model theory, as we have seen, implies that there will not necessarily be a natural kind to help in determining the broad content (nor even to be rigidly baptised in the first place), nor will there necessarily be a nomological “way the actual world could be” that matches any of the folk conceptual possibilities. It is in virtue of these sorts of semantic connections that propositions about water get to be true on many existing semantic accounts. Moreover, the next best natural kind—if there is one to be found—will likely fail to instantiate many of the properties that water is narrowly represented as having.

I have only been able to give a suggestive sketch of the behavioural model theory of perceptual representation and its rationales, and many of its details and implications will have to be further developed. Whatever implications the theory might have for science and for a semantics for natural languages, I think that the foregoing supports its viability as a naturalistic means for characterising perceptual events and the way these events affect language and reasoning. If the theory were accepted, it would thereby figure importantly in constraining any account of narrow mental and linguistic contents. Its focus on behavioural contribution is congenial to both the proper consideration of the natural selection of the perceptual systems, as well as to the synchronic use of perceptual activity by the system as a whole. By avoiding the presumption that perceptual states represent natural kinds, it presents an account of perceptual representation that is more naturalistically defensible than many rival accounts of mental content. Finally, the theory grounds representations in the intrinsic causal organisation of the system in a manner that gives them important explanatory roles. The theory thereby succeeds in providing the basis for a robust realism with respect to mental representations71 in the wake of accounts by realists that fall variously short of putting mental contents to good theoretical use.

71 The theory indeed provides a basis for robust realism, but recall what this amounts to in light of [EN] (§1.1).
References


Cambridge, MA: MIT Press.