Correspondence and Construction:

The Representational Theory of Mind and Internally-Driven Classificatory Schemes

Gabe Dupre

Keele University

# 1 Introduction

There is a tension at the heart of much contemporary work in philosophy of psychology— specifically, within *representational* theories of mind. On the one hand, the central insight of this tradition is that mental and behavioural processes are understood by appeal to mental representations: mental tokens which function as internal proxies for some aspect of the environment, on which behavioural interaction with the environment can depend. On the other, it has long been noted that many purported representations seriously distort, or even simply fabricate, those aspects of the environment they are alleged to represent. I will focus on the examples of colour vision and speech perception. At a minimum, this puts pressure on the explanatory goals of representationalism. Many representational theories explain behaviour with reference to *accurate* representation, but if we can seemingly function perfectly well with wildly inaccurate representations, the centrality of this strategy is threatened. At worst, this undermines the representationalist project itself, posing insuperable worries for any account that seeks to ground mental content in relations to the environment.

# 2 Representation as Correspondence

The core idea of the representational theory of mind is that mental states function to correspond to the extra-mental world. Having an internal state which represents the external world allows organisms to better navigate their environments, as they can make their behaviour sensitive to the properties of this internal state. Conditioning behaviour on such internal representations brings with it many benefits. For one thing, it increases the range of environmental properties behaviour can be sensitive to. Organisms’ sensory transducers are sensitive only to a small subset of the properties of the environment (wavelength, intensity, and polarisation of light, chemical structure of odorants, pressure on the skin, etc.). Interposing a system of representations between these sensory transducers and behavioural responses grants the latter sensitivity to non-transducible properties: I can run from predators, and towards food, rather than simply respond to light and sound. Further, representations provide stability. The ways the environment causally impinges on the body are hyper-sensitive to minor changes in the body, the environment, or their relation. Conditioning behaviour directly on such interactions would thus be likely detrimental. While the retinal projection from an object may shift radically as I move my head, my representation of the object can remain the same, allowing me to continue interacting with it. On a longer timescale, forming a representation of the spatial layout of a maze means that the next time it is encountered it can be run efficiently rather than painstakingly explored. And so on.

In these ways, mental representations can be thought of analogously to public representations, like maps. If I want to get from Shepherd’s Bush to Elephant & Castle, my best bet is to condition my behaviour on the spatial relationships schematically depicted in the Tube map. The map in this case serves as a proxy for the real-world connectivity relationships between the stations. The crucial information contained in the map (e.g. that I need to change at Oxford Circus) is not easily accessible without such an intermediate representation. And to the extent that the map is accurate, my behaviour will likely be successful (assuming normal service). The representational theory of mind thus views mental states as functioning similarly to these external representations: serving as stable stand-ins for the environmental properties of interest.[[1]](#footnote-1)

While details vary, in broad outline this view has been espoused by theorists of quite

different orientations:

* The term representation focuses attention on the image of the distal object in the symbolic structure which represents it. The analysis [. . .] presented here, focuses on the adequacy of constructing functions from the distal object to the behavior of the system, which works through the representations as an intermediate structure. (Newell, 1980, p. 176).
* A mental representation is a functioning isomorphism between a set of processes in the brain and a behaviorally important aspect of the world. (Gallistel, 1998, p. 13)
* To a first approximation, a given kind of animal comes to have an internal model of its world; that is, of its relevant-to-my-life-style world. (Churchland et al., 1994, p. 56)
* I have argued here that a particular type of representation, referring to patterns of activity that bear a systematic relationship to the structure of the external world and play a causal role in behavior, is fundamentally necessary for any intelligent organism. (Poldrack, 2020, p. 16)

As these quotes make clear, the representationalist strategy is widespread in philosophy and the cognitive sciences. However, on pain of vacuity, there must be constraints on when representations are profitably posited. I believe the following simplified schema (drawing primarily on aspects of Fodor (1986), Burge (2010), and Shea (2018)) identifies the crucial features of inquiry which justify such posits.

1. An organism is observed to display behavioural plasticity.
2. This plasticity cannot be adequately explained merely with reference to differences at the sensory periphery.
3. This plasticity can be adequately explained with reference to differences in distal features of the environment.

When these three conditions are met, we can reasonably conclude that the organism mentally represents the distal features their behaviour is sensitive to.

By condition 1, it is meant that an organism shows a range of different behaviours across different environments. Of course, there is vagueness in this statement (how big a range? How different must the environments be? etc.), but not I think in a way that undermines what is intended. This condition is relatively uncontroversial: I think every major account of representation takes the explanation of differential response to a changing environment as central. It is also very weak. Most organisms, including relatively behaviourally restricted ones such as plants and microbes, can be claimed to meet this condition, and even extreme non-representational theories (e.g. behaviourism) view behavioural flexibility as a central explanandum.[[2]](#footnote-2)

Condition 2 is more substantive. It ensures that we do not explain with appeal to representation that which can be explained solely as a matter of causal response to sensory information. Representational theories get their bite by positing an extra, intermediary layer between sensation and behaviour. A system which fails to meet condition 2 will thereby not call for such explanation. Phototropism in plants and chemotaxis in some microbes seem like paradigm cases here. An *E. coli* bacterium will display characteristically different movements depending on the chemical gradients of attractants and repellants in its local environment, but this behaviour is strictly conditional on the impact of these chemical stimuli on its receptors over short time scales. For this reason, there seems to be little call to view these organisms as representing, rather than merely sensing and responding to, chemical stimuli. One can think of this as an application of Morgan’s Canon, which warns against attributing “higher” psychological capacities to organisms when “lower” capacities are sufficient. Early defenders of the representationalist approach such as Chomsky (1959) and Fodor (1968) argued for representations precisely on the grounds that many human behaviours did meet condition 2, and thus were not susceptible to behaviourist-style explanation.

Paradigmatically, organisms will meet condition 2 when there is a “mismatch” between sensory input and behaviour. That is, either the same input causes different behaviours, or different kinds of input cause the same behaviours.[[3]](#footnote-3) This degree of freedom between sensation and behaviour is filled in by intervening representational systems. This is a nomological condition: when there are *laws* mapping sensation onto behaviour, there will be no benefit to positing representations.[[4]](#footnote-4) Condition 2 is thus necessary for positing representations, but it is not sufficient. There can be gaps between sensation and behaviour which are not best accounted for by representational systems. Most obviously, a system which acted randomly would meet conditions 1 and 2, displaying different behaviour in different contexts in a way not explicable with reference to sensory input. But the behaviour would be no more explicable with reference to representational systems.

For this reason, condition 3 is required as well. The behaviour of interest is *appropriate*, in light of the distal environment the organism is in. Objective properties of the environment, which outstrip the information conveyed to the organism in its sensory channels, explain why a certain behaviour was successful. We say that the creature represents *depth* because its behaviour is appropriately related to the depth of the objects in its environment, even though depth is conflated with, among other things, size in the information made available by its sensory systems.

When these three conditions are met, positing representations is explanatorily powerful. Representations interposed between sensation and behaviour allow for the complex causal pathways between the former and the latter. And appeal to distal features of the environment specifies what role this intermediate system plays and why playing this role is useful to the organism: the representations stand in for these distal features, and this is useful because conditioning behaviour on these features rationalises the organism’s actions.

This is, of course, an inference to the best explanation, and thus a fallible, ampliative inference.[[5]](#footnote-5) Specifically, positing representations generates predictions for neuro-biological accounts of the mind. Internal representations of distal states of affairs require physical vehicles capable of acquiring, storing, and making available information concerning whatever aspects of these distal properties downstream behaviour is sensitive to. While it is far from straightforward to figure out how such information is stored in the brain, it is clear that it must be, for the representationalist project to succeed, and some physical systems will be more plausibly suited for this than others. As always, our scientific theories must be evaluated for both their successes within their proprietary domains and also their fit with other disciplines.

These are then quite stringent conditions. However, there are ample cases in the literature of psychological systems meeting them. Some of the clearest come from visual perception of “physical” properties (such as distance, size, and motion), and from the navigatory capacities of various creatures.

One of the basic tasks for any visual system is locating local objects in space, relative to the location of the organism. One way to determine the location of an object in 3-dimensional space, relative to a perceiver, is by identifying a line between object and perceiver and then determining the distance along this line between the two. The former is relatively straightforward: that light travels in a straight line means that this information is generally extractable from the retinal image. However, the latter is complex. There are no laws that guarantee that a certain retinal image corresponds to an object of a certain distance from the viewer. That this is so is evidenced by the conflation of size and distance in retinal projections: big objects far away can project retinal images of the same size and shape as small objects nearby. For these reasons, visual systems have developed multiple complex and subtle strategies for inferring distance from retinal cues. One such strategy is reliance on *texture gradients*. Many distal scenes contain surfaces featuring repeated patterns. Checkerboard-tiled floors are paradigms here, but similar examples can be found in natural scenes, such as the roughly evenly sized pebbles covering the ground on a beach. When such surfaces extend away from the perceiver, despite their intrinsic regularity, they project non-repeating patterns onto the viewer’s retina. “Units” of the pattern (individual tiles, pebbles, blades of grass) near the perceiver project distinct shapes, with clear demarcation between them, whereas those far away blur together; nearby units will project bigger images compared to those far away; and so on. Gradients of size, distinctness, distortion, and other retinal features can thus be reliably correlated with regularly patterned surfaces extending away into the distance. The visual system can thus rely on this correlation, and infer that the distal scene in front of them contains a regular surface with parts at varying distances, rather than other alternatives physically consistent with this retinal image, such as a 2-dimensional surface perpendicular to the viewer’s line of sight, with large, well-defined units of a pattern at the bottom and smaller, less-defined units at the top. This preference for one distal scene over another, despite their both conforming equally well to the proximal projection, is achieved due to what Tyler Burge calls “formation principles”: psychological rules which govern the transition from proximal stimulation to perceptual representation. The rule that generates a representation as of a surface extending away from the viewer on the basis of texture gradients in the retinal image provides a paradigmatic example of a formation principle.[[6]](#footnote-6) Similar stories can be told for other kinds of environmental property, such as size or motion, in terms of reliable (but imperfect) correlations between distal properties and retinal projections, which are reflected in formation principles.

Crucially, while such formation principles allow the possibility of error (as when well-constructed paintings give an “illusion of depth” through just such depth cues as texture gradient), the correlations they depend on are generally environmentally valid, and so in most cases they enable us to accurately perceive our environment. The reliance on such principles thus enables internal states to accurately report properties (e.g. depth) of distal objects. These accurate reports can then guide action, as when we successfully manage to grab a nearby object. Thus, depth perception seems to perfectly fit our schema for representational explanation. Proximal stimuli seem unable to explain why we reached for the object at the depth that we did, given that they are consistent with objects at different depths. But the distal property (i.e. the actual depth of the object) does seem to explain why we reached where we did, as we reached successfully. So, this distal property must be represented by some internal state suitable for guiding behaviour.

Another relatively well-understood case comes from accounts of animal navigation. Again in such cases, we see the computation of (generally) accurate information about distal features, partly on the basis of variable proximal stimulation. Gallistel (1998) provides a comprehensive review of the capacities of ants, bees, and wasps to generate internal “maps” of their surroundings, incorporating both egocentric and allocentric information (the latter identified relative to the movements of the Sun) about the relative distances and directions of landmarks and locations of interest such as hives/nests and food sources. The information contained in such a map can then be used to solve novel navigatory tasks, such as returning home from novel locations, and orienting towards landmarks seen from novel directions. These insects rely on processing principles reflecting geometric laws, analogous to those used by human navigators when determining location on a map from the angles generated between their position and three landmarks.

In these and related cases, it is relatively straightforward to identify the environmental property that such systems serve to identify, and on which subsequent processing then comes to rely. These cases thus provide paradigmatic examples of the representationalist schema. The bee identifying its location on a map using a “three-point fix” cannot, of course, apply geometric inference rules to the angles in fact found in the environment, so must generate internal symbols that correspond to them and compute its location on the basis of these symbols. To the extent that such symbols correctly correspond to the angles in the environment they represent, the bee’s subsequent navigations will be successful. As these environmental correlates are objective, i.e. identifiable without reference to the bee’s internal states, these correlations can be verified. The internal, computational story, appealing to re-usable mental symbols being manipulated in lawlike ways, and the representational story, appealing to the distal properties such symbols represent, thus go hand-in-hand.

These systems, then, will serve for me as paradigmatic examples of the standard story about mental representation as serving to provide an internal proxy for a distal property, object, or state of affairs, on which organismic behaviour can be conditioned. If I generate an internal state which accurately corresponds to the distance between my body and my coffee cup, I am able to pick the cup up safely. If I misjudge this, I am liable to knock it over. Likewise, we can explain how the ant makes it home from a novel location by appealing to the accuracy of its internal map and its self-placement on this map. Thus, these cases provide paradigmatic examples of what Shea (2018) calls the “explanatory grammar” of representational explanation, according to which successful behaviour is explained with reference to accurate representation, and unsuccessful behaviour is explained with reference to inaccurate representation.

**3 Representation Without Correspondence?**

Despite these clear cases, and the fairly wide-spread view that representations serve as proxies for distal features, and are explanatory precisely in virtue of this, various theorists have pointed to a range of cases which don’t seem to fit this mold. In particular, there are cases which are commonly described as representations by both philosophers and psychologists, but which do not seem to correspond to anything in the creature’s environment. The point is made forcefully by Akins (1996) who says that “In general, sensory systems both make and ignore distinctions in nature when it suits the organism’s motor needs.” (p. 369).

To generate a difficulty for representational theories, we need to find cases which seem to be representational, in that behaviour seems to be conditioned on an internal symbol which is not explicable solely as a response to proximal stimulation, but for which there are no plausible external states for them to represent. This is made tricky by substantial disagreements about which sensory systems should be viewed as generating representations. For example, Barwich (2019) argues compellingly that olfactory sensation should not be understood as functioning to internally reproduce environmental states. Whether this is a worry for an account of *representation*, however, depends on whether we agree with Lycan (2014) that olfaction is representational, or side with Barwich and Burge (2010) in denying this status to the chemical senses. Similarly, Akins (1996) argued that thermoception radically distorts environmental signals, on the basis of biological relevance to the organism, thus posing problems for an intentional theory of the sort outlined above for our detection of environmental temperature. However, again Burge (2010) has argued that what this shows is not that representation is not a matter of reliably standing in for some distal state, but rather that thermoception is not representational.

At the other end of the scale, “higher” forms of cognition appear to provide novel ways of determining mental content. Our abilities to think, and talk, about fictional entities, and to develop abstract scientific theories, for example, outstrip the explanatory resources of the above account of representation. Following Burge (2010) and Shea (2018), I shall thus restrict my attention to representational systems that do not depend for their functioning on complex propositional or linguistic reasoning.

For these reasons, I will focus my attention on two problem cases that are both (i) widely agreed to be representational, and which (ii) display many of the characteristics of the genuinely representational systems discussed in the previous section, which (iii) do not depend on conscious, non-modular, reasoning, but for which (iv) there seem to be no external entities, states, or processes which they function to represent: namely, perception of colour and of speech. My discussion will be quite brief and highly simplified, serving mainly to give a flavour of the ways in which the mind creates and imposes structure of its own on stimuli, rather than seeking to reproduce the properties of these stimuli.

While I will get into some detail with these examples, I do not mean to suggest that these are the sole cases which generate the issues discussed in this paper. I believe similar arguments could be made with appeal to a variety of other psychological capacities. Further, even if all such cases can be resolved in ways amenable to the representationalist style of explanation adumbrated in the previous section, that would be a significant empirical result. As I will discuss them, colour and speech perception represent the empirical possibility that psychological systems could function internally *as if* they were tracking environmental properties, without such properties being present.

Both colour and speech perception pose various distinct puzzles for representationalist theorising. The general point is that in perceiving colour and speech, the mind creates its own structure. On one standard way of talking, perceptual categories of colour and speech are mental *constructs*. Classifying something as red, or as a /b/, is not a matter of determining which category the external stimulus antecedently exemplifies, but rather of imposing an internally-driven classificatory scheme onto the stimulus.

The first thing to notice is that both colour and speech exemplify *categorical perception*. This is the phenomenon of perception imposing a discrete grouping onto continuously variable stimuli, and treating those members within a given grouping as more closely related than they are to those outside of the grouping, even if the “objective” distance between the in-group and out-group stimuli is equal to or even less than that between in-group stimuli. For example, consider a beam of white light being dispersed by an optical prism. The prism will produce a spectrum of rays of continuously varying wavelength. However, when we look at this spectrum, it will not look continuously varying to us. Rather, it will look to be composed of a relatively small number of bands of colour, with blurry edges between them. Perceived similarities between locations along the spectrum will not be determined by “objective” similarities of wavelength, but rather by their place in these colour bands. So a blueish green will look more similar to a yellowish green than it will to a greenish blue, due to the former both being classified as *green*, even if the wavelengths of light responsible for these percepts are more similar in the latter case than the former.

The same thing is observed in speech perception. Most famously, plosive, or stop, consonants like /b/ and /p/, generated by the blockage and then release of air in the oral cavity, may differ continuously in their Voice-Onset-Time (VOT), the time between the start of airflow after the blockage of the airway and the vibration of the vocal cords. However, despite this continuous variation in their articulation, these sounds are not perceived as forming a spectrum, but as discrete categories. For example, bilabial plosives, formed by stopping airflow by closing both lips together, are differentiated according to VOT: those with a VOT of 30ms or less tend to be perceived as /b/, while those with VOT of over 30ms are perceived as /p/. A consonant with a VOT of 28ms will typically be perceived as more similar to one with a VOT of 15ms, due to both being /b/s, than to one with a VOT of 32ms, despite the latter being more “objectively” similar. So, we could arrange a series of stimuli ordered by VOT, with arbitrarily small gaps between each instance. From the perspective of articulation, we have a roughly continuous series, with VOT increasing steadily. However, perceptually, what we hear is not a /b/ sound gradually changing into a /p/ sound, with some intermediate cases along the way. Rather, we hear this as a series of definite /b/ sounds, and then suddenly a series of definite /p/ sounds. This sudden transition must thus be imposed on the stimulus by the mind, rather than correspond to some pre-existing environmental change.[[7]](#footnote-7)

Categorical perception provides an initial illustration of the difficulties posed for representationalist accounts. The categories, and thus the perceived “similarity space” imposed by our perceptual systems, do not seem to correspond to any objective feature of the stimulus. Unlike in the case of, say, depth perception, where we could look to the environment to see whether a percept was accurate or not, in cases of categorical perception we cannot do this. The physical facts about the environment, wavelength or VOT, do not determine whether a stimulus will be perceived as greenor as a /b/. Information about the internal classificatory system, about where these lines are drawn, is needed in addition.

While illustrative, I don’t view categorical perception on its own as damning. While the categories seem to be internally determined, that some stimulus falls into one category or another could still be viewed as an objective fact about the environment, which perceptual systems function to track. *Merely* imposing a classificatory scheme on environmental variation may be compatible with the general representationalist account identified above. What will pose a problem is if such classification schemes go beyond merely dividing up the objective stimuli and distort the objective relationships, or even invent classifications *de novo* with minimal input from the environment. Discussing examples of just this phenomenon will require presenting a bit more of the details of colour and speech perception.

“Colour vision” refers to the collection of capacities of organisms to differentially respond to different wavelengths of light. Wavelength provides a valuable source of information concerning the organism’s environment. However, making use of this information typically requires significant processing due to the fact that wavelength conflates a variety of different environmental sources. Most significantly, wavelength reflected from a surface is a product of both the surface reflectance properties of the object and the wavelength of the light illuminating it. A white wall in red lighting may thus reflect light with the same wavelength as a red wall in white lighting. It has traditionally been assumed that the purpose of colour vision was to extract information about the colour (i.e. surface reflectance) of distal objects, and thus the goal is to control for variations in such confounding influences as variation in illuminants. Such a picture fits nicely with the view of representation as correspondence detailed above.

However, much work in colour science tells against this view. Many features of colour vision would be unexpected on the basis solely of examination of the reflectance properties of distal objects. For example, human colour vision identifies four colours as “unitary”, i.e. as not decomposable into mixtures of other colours. So, while orange and purple *look* to us like mixtures of red and yellow and red and blue, respectively, red, green, yellow, and blue look to be basic.[[8]](#footnote-8) Our judgements of colour similarity and difference are strongly influenced by such classifications. While the neurobiology and psychology are complex, it is widely accepted that this classification is explained at least partially with reference to internal features of the trichromat visual system, especially the post-receptoral “opponent coding” system which is traditionally understood to respond to the ratios of red to green and blue to yellow light. That is to say that the representational system used by the colour vision system is not explained with reference to the properties it tracks in the environment, but rather by the physical channels involved in sensation.

Of course, providing a neurobiological explanation of our visual classification schemes doesn’t *preclude* a correspondence-based account. Indeed, one might be tempted to think that standard “methodological adaptationist” (Godfrey-Smith, 2001) assumptions about the evolution of mental capacities might tie them together nicely: our visual systems make the distinctions they do because latching onto certain environmental properties was beneficial for our ancestors. However, even if we accept this adaptationist explanatory scheme, we cannot simply infer that the function of such mental capacities is to correspond to environmental distinctions. One can accept that colour vision is useful without accepting that it is useful for tracking, as do Akins & Hahn (2014) and Chirimuuta (2015). And given the above arguments that colour vision doesn’t seem to involve tracking distinctions in our present environment, we seem to have strong reasons to deny that any phylogenetic story will be able to save the correspondence-based approach to colour vision.

A related worry for representationalist approaches comes from metamers: surfaces with radically different reflectances which, in a particular context (e.g. under specific illumination, or viewed from a specific angle), appear to be identically coloured to an observer. Metamers raise worries for any account which views colour vision as functioning to track environmental properties, as it seems these quite distinct environmental properties (surface reflectances) are generating identical representations (as of the same colour). Unless the representationalist is willing to say that in each class of metamers, all but one systematically leads to inaccurate perception, this suggests that the represented category is one that does not correspond to any objective, or mind-independent, feature of the environment.

Turning now to speech perception, we see the same sorts of problems arising, wherein the categories utilised by the system seem to be internally-driven, rather than reflective of any objective environmental attributes. Speech perception involves categorising a speech stimulus with respect to a hierarchy of linguistic properties: sounds must be classified phonetically and phonologically; sequences of phonemes must be grouped into syllables, morphemes, and words; words must be grouped syntactically into phrases, and semantic values must be assigned at both the morphemic and syntactic level. At every stage of this process, there are vast gaps between the imposed mental classification and any properties of the stimulus.

Consider first the lowest levels of identifying phonological properties of an uttered string. Each language determines a set of phonemes, usually around a few dozen, out of which the legitimate sounds of the language can be composed. Identifying the words one hears depends on identifying the phonemes composing them (*did she just say ‘bin’ or ‘pin’?*). Further, these phonemic inventories vary from language to language (e.g. dental fricatives, as in English ‘*this’* and ‘*brother’*, are absent in French), and so must depend on early exposure to native language. So both in use and development, the phonemic inventory for a given language must be, in some sense, identified by the child on the basis of environmental stimuli. This is, however, a notoriously difficult computational problem due to the very messy relationship between phonology and physical stimuli. The phonological perception system must be able to abstract from variation due to factors such as age, sex, (some aspects of) dialect, conversational focus and carefulness, etc., and group together the same phonemes despite significant variation in physical properties of the stimulus.

One well-studied example of this sort of computation stems from the fact that the anatomical features of the vocal apparatus impose constraints on how phonological forms can be efficiently produced. Different phonemes are (generally) individuated by the way they are produced by the tongue, vocal cords, and other parts of the oral cavity. In the rapid process of speech, however, the different features of such productions can be run together, in often predictable ways. For example, in English, when (what would normally be) a dental plosive (/d/) is immediately followed by a bilabial consonant (e.g. /m/, /p/, or /b/), it is typically transformed into a bilabial plosive. In most spoken English (i.e. as long as the speaker isn’t taking special care to speak slowly and enunciate), the final consonant of ‘could’, in ‘I could make more’, ‘He could pass the salt’, and ‘You could buy some’ will be realised not as a /d/ as it is written and would be produced in default contexts, but as a /b/. It is difficult to quickly transition from having one’s lips apart as needed for the production of /d/ to pursing them as needed for subsequent bilabial consonants, and so the system uses a “shortcut” and simply turns the /d/ into a /b/ to make production easier. This process is called assimilation.

What matters for our purposes is that assimilation appears (see e.g. Mitterer & Blomert (2003)) to be “corrected for” early in speech perception. That is, despite the different ways that ‘could’ is pronounced in different contexts (with or without assimilation), it is perceived as the same word, with the same, default, final consonant. Whether the stimulus is /kΩd/ or /kΩb/, it is *heard* as if it is the former. This mental correction of the assimilation imposed by speech mechanics provides a case of an internal classification system being imposed on an external stimulus, rather than the perceptual system functioning to correctly reproduce what is in the environment.

Any perception in a noisy environment is liable to generate cases of misrepresentation. Indeed, that such misrepresentation is possible has long been one of the central desiderata for a philosophical theory of mental content (Fodor 1990). However, assimilation does not seem to be merely another example of misrepresentation. On standard, hierarchical, models of speech perception, phonological representation of the stimulus feeds into lexical identification (which then feeds into syntactic and semantic parsing). Assimilation generates a tension between these two processing pressures: if we failed to correct for assimilation, and heard the spoken word as /kΩb/, this would impede our ability to identify it as a token of the word ‘could’. So, unlike paradigmatic instances of misrepresentation which reflect (perhaps necessary) computational shortcomings in our perceptual systems, in cases like assimilation what we see is a competition between the goal of representational accuracy (which would favour correctly representing the produced /b/ sound)and the internal usefulness of classification (which favours incorrectly representing the stimulus so as to make lexical identification easier), and in which the latter virtue wins out. Thus again we have a case in which the functioning of the system does not seem to be a matter of tracking environmental properties, but instead of classifying according to an internally-motivated scheme.

One final example comes from the process of syllabification, which groups phonemes together to form pronounceable sounds. As with phonological inventories, syllabification differs from language to language. A major driver of inter-linguistic variation in syllabification is phonotactics, the rules governing which sequences of phonemes can appear in which contexts. For example, languages typically place more constraints on the presence of consonant clusters at the beginning of a syllable than at the end. In English, many sounds which are fine syllable-finally are prohibited syllable-initially, such as /rt/ and /ts/ (‘cart’ and ‘bats’, but not \*‘rtac’ or \*‘tsab’).[[9]](#footnote-9) In perception, the string of phonemes must be divided up to form syllables. Phonotactics plays a central role here: syllables can’t be formed which would violate phonotactic rules of this sort. And this leads to inter-linguistic variation in how the same speech stimulus will be perceived. In Japanese, unlike English, /ts/ is legitimate syllable-initially (in the “onset”).[[10]](#footnote-10) Further, consonants are generally (with a few exceptions) precluded from appearing at the end of a syllable (in the “coda”), so that Japanese syllables almost always end with a vowel. Thus, certain words will be syllabified quite differently by native speakers of English and Japanese. The name of the car company, Mitsubishi, for example will be syllabified as Mit-su-bi-shi by English speakers, due to the prohibition of syllable-initial /ts/, whereas Japanese speakers will hear it as Mi-tsu-bi-shi, with the syllable-initial /ts/ allowed, and indeed required to prevent a prohibited syllable-final /t/.

Thus we again have a case wherein how an environmental stimulus is characterised seems to involve imposing an internally-driven classificatory scheme onto it, rather than seeking to identify its mind-independent features. That one and the same stimulus (e.g. a spoken token of the word ‘Mitsubishi’) can be classified quite differently by different speakers, without any pressing reason to determine that one of them is *misrepresenting*, problematises the idea that mental representation is a matter of reproducing environmental properties.

Note that these cases are arising in the lower levels of speech perception, wherein what is being classified does seem, intuitively at least, relatively close to the physical properties of the stimulus: sounds and sound-groupings. The problem of course gets much worse as we turn to higher levels of perception, such as syntax and semantics. Syntactic theory says that ‘the guests are eager to eat’ and ‘the guests are easy to eat’ have radically different grammatical structures, and psycholinguistic theories of human parsing provide theories of how such structures are generated as representations of speech stimuli. But whatever this difference consists in, it doesn’t seem to be an independently identifiable (i.e. identifiable without actually using the very perceptual categories under investigation) property of the stimulus in the way that relative spatial locations were in the examples discussed above. It certainly won’t be identified through spectrographic analysis. Likewise for the quite different semantic properties of such examples.[[11]](#footnote-11)

Despite these stark differences between colour and speech perception, on the one hand, and the perception of physical features of the environment and animal navigation, on the other, from a particular perspective we can note significant similarities. These differences consist exclusively in the relationships between these symbols and the outside world. These differences disappear when we view these systems from an “internalist” perspective, focusing only on the ways that internal states are formed, manipulated, stored, and relied on in controlling action. The “symbolic system” approach discussed in the previous section applies equally well to paradigmatic work in colour perception and psycholinguistics, involving the lawlike generation of internal tokens on the basis of these tokens’ membership in various equivalence classes, and subsequent processing determined by which such class they are in.

Before concluding this section it is worth briefly driving home the point that these systems cannot easily be dismissed as “merely sensory”, as opposed to full-blown representational systems. Burge (2010) has argued that perceptual constancy mechanisms should be viewed as the defining feature of representational systems: systems with constancies represent, sensory systems without constancies do not (in Burge’s terms, they merely “register information”). Constancy mechanisms, for Burge, are ‘capacities to represent environmental attributes, or environmental particulars, as the same, despite radically different proximal stimulations’ (2010, p.114). It is near-uncontroversial that this applies to both speech and colour perception (indeed, Burge discusses colour constancies alongside other constancy mechanisms). We see objects as having stable colours despite changing illumination or distance from the perceiver. And we hear speech sounds as consisting of the same phonemes despite varying features of the speaker. Relatedly, it is clear that what is involved in speech and colour perception goes far beyond mere averaging, aggregation, etc. of properties of the sensory stimulus. What these cases bring out is that the existence of a constancy is neutral on whether the environmental particular in question is being stably represented as having a property that it actually has or not. Size constancies enable us to accurately gauge the size of an object, despite the changing projection it produces on our retina. Colour constancy, on the other hand, enables us to stably impose internally-driven colour categories on distal stimuli, despite the variable ways such stimuli causally interact with our sensory systems. But this does not seem to require that the imposed category functions to track a genuine property of the stimulus. But if even someone as stingy with their attributions of “representational” as Burge seems forced to accept that speech and colour perception are representational systems, the fact that these systems *impose* their classificatory scheme on environmental stimuli, rather than functioning to track antecedently existent environmental categories, seems to suggest a deep conflict with the above-described account of when and why representations are to be posited.

Colour and speech perception, then, pose a stark worry for representationalist accounts of mind. On the one hand, these accounts are primarily motivated by the thought that internal proxies for inaccessible features of the environment play central roles in explaining organismic behaviour within such an environment. Representational contents are posited essentially as part of this explanatory project. But in the cases described here, there seems to be no aspect of the environment that these representations function to reproduce, or correspond to. The properties of colour vision, such as colour similarity and difference, are not properties of distal objects’ surfaces. And properties of heard speech do not, generally, track properties of speech stimuli, but internally-driven features of linguistic competence. However, these are often given as paradigms of representationalist psychology. This tension between the explanatory schema invoked in philosophical accounts of scientific methodology and between the empirical facts of the science must be resolved somehow.

# 4 Possible Responses

While my main goal in this paper is to identify a foundational worry for standard representational theories of mind, I will close by briefly outlining the various strategies one could take to resolve the problem outlined above. I will discuss them separately, but it is likely that a solution will involve elements of each.

Firstly, there are attempts in the literature to show how both colour and linguistic properties are genuinely found in our environments. Perhaps most famously, Byrne & Hilbert (2003) argue that colours are simply surface reflectances, colour vision functions to represent surface reflectances, and that the representations formed are often accurate. If this can be made to work, despite the criticism raised in e.g. Hardin (2003), then it would be possible to assimilate colour to the physical properties for which representational theories worked so well. A similar account could be offered for language. Pereplyotchik (2017), for example, defends a possible combination of Devitt’s (2006) view that public linguistic symbols have the phonological and syntactic properties posited by linguistic theories in virtue of conventions among speakers with a representational theory of speech perception. Here too, these proposals have received strong pushback—see for example Collins (2008) and Rey (2006) for worries about Devitt’s claims about the ontology of linguistics. If these respective worries can be overcome, this would provide a first step towards solving the puzzle discussed above, and showing how the standard representation-as-correspondence account could apply to colour and language.

While these strategies are available, I wouldn’t bet on them. For whatever it’s worth, within philosophy, they are not hugely popular, and within vision science and linguistics they seem to be vanishingly rare. Perhaps the central reason is that even if we can make ontological sense of an environment populated with reds and blues and nouns and verbs, it has often seemed that doing so is a sort-of metaphysical third-wheel. That is, the ontology does not seem to play a central role in the explanatory projects of the sciences. If the proposed identification of colours with reflectances failed, vision science would not, it seems, have to change anything. Likewise, the success of psycholinguistics does not seem to turn on the question of whether there are unarticulated constituents in our environments. As the motivation for positing mental representations, at least in the tradition I have been focusing on, has centrally been an explanatory one, it would be an uneasy marriage between this general approach to mind and these apparently non-explanatory ontological doctrines. Further, as mentioned, speech and colour are illustrative examples, but there is no reason to think they are the only problem cases. Any account which aims to solve the problem by populating the environment with suitable targets for representation would thus have to generalise to cover all such cases.

Further, I said that this would be a key *first step* in resolving the problem. But more would be needed. In the terms of Quine (1951), this would account for the *ontology* of our theories of colour and speech perception, but not necessarily for their *ideology*. That is, it could ground claims that colour and speech perception function to identify examples of colour and speech in the environment, but it may leave unchanged the central point of my argument: that the categories involved in classifying such stimuli are internally generated. Take the case of colour, and allow with Byrne and Hilbert that surface reflectances are colours[[12]](#footnote-12). From this fact alone, it doesn’t follow that when we classify some environmental colour as *red*, we are thereby identifying a property that it has independently of our so classifying it. For that, we would also need to show that there is some mind-independent reality to the property *redness*, instantiated by objects with quite different surface reflectances. But, as Byrne and Hilbert allow, such a property ‘will be quite uninteresting from the point of view of physics or any other branch of science unconcerned with the reactions of human perceivers.’ (2003, p. 11). But this seems to be just another way of saying that such a classificatory scheme is internally determined, not a matter of identifying or tracking a pre-existing environmental distinction. Likewise with linguistic properties: saying that there are linguistic properties in the environment is one thing, and saying that our perceptual classification of them involves placing them into categories they instantiate independently of such a classification is another. For these reasons, I will assume that the problem above is a real one, and will turn to more radical approaches to resolving it.

One approach starts from the observation that misrepresentation is a central feature of any account of representation. From a theoretical perspective, it is often recognised that there can be no representation without the possibility of misrepresentation (Fodor 1990). And more prosaically, it is a commonplace that many of our representational states are misguided in some way or other, from perceptual illusions to cognitive delusions. One possibility for resolving the above worries then amounts to viewing perception of speech and colour as simply misrepresentation on a grand scale. Mendelovici (2013) argues for this with respect to colour, and Rey (2008, 2020) applies it to both of the cases I discuss. On this view, mental representations do indeed function to correspond to environmental properties, but that provides no guarantee that they succeed in so corresponding. So just as illusions of depth provide no objection to accounts of depth perception as representational, we should not view the fact that there are no colours or linguistic properties in our environments as reason to doubt that our perceptual systems represent these properties.

The main obstacle to a view of this kind is that it is in need of a meta-semantic account of what makes these representations the representations they are. The standard representational theory of mind discussed above provides a neat answer here: mental states represent those aspects of the environment which cause them, and to which behaviour is sensitive. Depth perception represents depth because different representations are formed in response to interaction with objects at different depths and it enables creatures to interact successfully with objects at different depths. Once it is admitted that there are no colours or linguistic properties in the environment, this opens up the question of why we view these properties as the represented contents, rather than some others. As Stalnaker (1989) argues compellingly, it is often not possible to individuate a mental content other than with respect to its causal antecedents. This may be compatible with occasional misrepresentation: representations of depth typically correspond to environmental depth, which suffices to ground their representational content, even if on certain occasions this correspondence fails. But it is much more difficult to see how it is compatible with *across-the-board* misrepresentation as we see in the cases described above.

A closely related concern is that such accounts seem to undermine core motivations for representational theories. As noted above, Shea (2018) points out that one central role for representational content in cognitive science is in licensing a particular “explanatory grammar”: ‘Representing correctly explains successful behaviour and misrepresentation explains failure.’ (p. 28). If representations of colour and speech are inevitably misrepresentations, then such an explanatory strategy cannot apply. Representational explanations paradigmatically function to show how an organism can engage successfully with its environment, but if my hallucinatory representations of the banana in front of me as blue are no less accurate than would be representations of it as yellow, what explanatory purchase does this content have?[[13]](#footnote-13)

Rey (2020) proposes a dispositionalist account here: representations of colour/speech have the content they do because *were* we to interact with genuine colours or Verb Phrases, our perceptual systems would track them, even if we never in fact encounter such things and so all of our representations of them are misrepresentations. I believe such an account similarly suffers from the above-mentioned worry that it is *actual* behavioural success and failure that we want our cognitive theories to explain, and any account on which our representations are uniformly inaccurate will not be equipped to provide this. A further problem is that no particular reason to believe such a counterfactual is given. Our perceptual systems are designed to respond to the physical stimuli they in fact encounter and have encountered over evolutionary time. It is thus not clear that they would recognise a genuine instance of redness or of a controlled unarticulated subject if they were to encounter them (if such objects/properties can even be made sense of). Given that Rey admits that we don’t encounter such entities, it is not clear why he is entitled to make claims about what we would perceive if we did.

Other sorts of metasemantic accounts could step in here, providing an alternative solution. While the representation-as-correspondence approach is fairly dominant within the cognitive sciences, within philosophy there is a significant tradition viewing representational content as determined in other ways. One widely discussed alternative is the inferentialist tradition, according to which the meaning of a symbol is determined primarily by its relations to other symbols within a broader representational system. In different ways, this idea has been developed by Loar (1981), Block (1986), Greenberg & Harman (2008), and Brandom (1994) (although see Fodor & Lepore (2007, 1992) for criticism). Representing a stimulus as *red*, or as a *rounded vowel*, on this view, would be a matter just of applying to it a representational token which plays a specified role within the overall perceptual systems of colour or speech. Such an account again faces the problem that it fails to explain why certain actions are successful and others are unsuccessful, as this seems to rely on correspondence, or lack thereof, between environmental and internal states. More generally, one could worry that such approaches are unsuited for behavioural explanation on the grounds that content is simply an abstraction over the causally significant mental symbols and processes. On the standard view, a state’s being an edge-representation depends on its corresponding to environmental edges (Burge 1986), and appealing to this mind-world relationship explains why having such internal states would be useful to an organism. But this style of explanation is unavailable to the inferentialist. If what makes something an edge-representation is simply standing in causal relationships to other sorts of representations, it doesn’t seem that classifying something as such a representation adds anything of explanatory value to the pre-existing causal/syntactic story. So one major challenge for the view that mental content can be grounded other than with reference to correspondence is to show how such content is explanatorily pertinent to cognitive and behavioural science.

Of course, there are also other metasemantic options one could take, complicating the relations between content, environment, and computational role.[[14]](#footnote-14) But I hope what I have argued above is enough to show that simply moving away from the content-as-correspondence view is not sufficient to avoid these worries. That there must be at least a plausible story, if not a fleshed-out theory, of what makes representations of colour and speech contentful, and that such a story must contribute to the explanatory goals of the cognitive sciences, together provide powerful constraints on a theory of mental content. So powerful that to my knowledge, no proposal in the literature has yet managed to meet them.

Given these worries about a suitably general and explanatory representational theory of mind, one could take these cases as demonstrations that this is simply the wrong way to think about the mind. That is, one could argue that colour and speech are just instances of the larger, “internalist”, idea that the mind should be studied on its own terms, rather than as relating to or reflecting its environment. This idea has been prominently defended in philosophy by Stich (1983), who argues for a “syntactic theory of mind”, according to which mental states are individuated by their computational roles, not by their representational contents. Within the cognitive sciences, Chomsky (2000) has consistently argued, centrally from the linguistic case discussed above, that one’s environment can play no role in individuating mental states, and thus that notions of representational content in the robust sense endorsed by the standard account should be rejected. Collins (2014) develops an “algebraic” approach along these lines.

If taken as general accounts of the mind, the standard worry for such approaches is that the assumption that we can provide equally good explanations of behaviour without reference to the environment is simply mistaken. As Pylyshyn (1984), Fodor (1987), Burge (2003), and others have argued forcefully, genuine cognitive-scientific explanations take patterns of mind-world interaction as both explanantia and explananda. Further, certain philosophically significant properties of mental life and behaviour, such as “objectivity”—drawing a boundary between the world and the way the world appears to us (Burge)—and rationality—explaining how our behaviour is appropriately responsive to our environmental conditions (Newell)—seem constituted by relationships between internal states and the surrounding environment. Any account of the mind which abstracts entirely from relations to the environment seems intrinsically unable to shed light on these phenomena.

To address this worry, the internalist position has more recently been developed in a series of papers by Egan (1989, 1995, 2009, 2014, 2018), who follows Stich and Chomsky in viewing mental states as individuated by their internal, computational role, while allowing that relations to the environment may play an explanatory, but non-individuating role. Rescorla (2017), however, argues that, in actual practice, empirical psychology individuates mental states precisely by their representational roles, and not, in general, by any syntactic categories that they may instantiate.

If these worries can be overcome, one nice feature of Egan’s proposal is that it allows room for a pluralism about cognitive explanation. If mental states have their representational properties only contingently, if at all, we can allow that these properties play an essential role in explaining cognitive capacities and behaviour in the cases in which correspondence does play a role, such as spatial perception. That some such connections to objective aspects of the environment are featured in our descriptions and explanations of the mind and behaviour can thus provide the raw materials for accounts of rationality mentioned above. But, we may still view such states as members of a broader computational kind which contains also states like colour and speech “representations”, for which there seems to be no possibility of correspondence.[[15]](#footnote-15)

Such a pluralism would take seriously Marr’s (1982) observation that a computational theory of any aspect of mind must tell us both *what* is computed, and also *why* it is this computation rather than some other that is found in a given psychological system (pp. 22–23). What the cases above then point to is the fact that answers to this latter question can take forms other than those Marr described, which involved computations for extracting distal information from an impoverished proximal stimulus. Much work in psychology provides alternative styles of explanation, but within the same information processing framework. For example, Averill (2005) claims that opponent-coding models of colour vision can explain why we classify colours as we do, despite such classifications not tracking any external properties. Chirimuuta (2015, ch.3) points out that Averill is here relying on a scientific conjecture, rather than accepted results, but for my purposes the possibility of such a model is sufficient. If the underlying neurobiology can provide an explanation for the ways that information is processed without appeal to estimating distal properties, then the pluralism I am here suggesting could go some way towards resolving the problem raised by representations without correspondence.

Alongside these explanations from neurobiological constraints, functional, but not correspondence-based, accounts of colour vision have been provided by Akins & Hahn (2014) and Chirimuuta (2015). On these views, colour vision is functional, but its function is not to reproduce external colour. Rather, colour vision provides us with information about a whole host of other environmentally useful properties, such as object boundaries and distance.

One sees such alternative models of explanations in linguistics as well, such as Chomsky’s (2001) explanation for wh-island effects which appeals to computational efficiency.[[16]](#footnote-16) Such an account explains why we perceive certain stimuli as having the linguistic properties they have (e.g. grammaticality, or the availability of a given reading) not with reference to them actually having such properties, but instead by claiming that representational systems which allowed for different grammatical structures (e.g. with long-distance wh-movement) would be too computationally costly.

By carefully distinguishing the computational story concerning what role mental symbols play within a cognitive system from the explanatory story concerning why such symbols are found with the properties they have, we can thus highlight the reasons why we felt compelled to group colour and speech with paradigmatic cases of representation as correspondence as well as the key differences between them.

While there is of course much more to say here, my aim has been to reasonably clearly identify the major strategies for resolving the problem raised by representation without correspondence. The hope is that some, likely pluralistic, combination of the following will provide the tools for resolving the problem: i) identifying suitable targets for apparently constructive representational systems in the environment, ii) providing a metasemantics for such systems which allows for widespread misrepresentation, iii) providing non-representational explanations for such capacities, and iv) providing an account of representational content which does not rely on correspondence. Such a solution would ideally retain the explanatory successes of the standard, correspondence-based approach, while showing how to generalise to cases like colour and speech.

# 5 Conclusion

In this paper I have argued that there is a deep tension at the heart of contemporary representational approaches to philosophy of mind and cognitive science. On the one hand, the explanatory apparatus of these disciplines seems to essentially involve appeal to correspondences between mind and environment. On the other, paradigmatic representational systems, such as colour vision and speech perception, seem to lack any suitable environmental correlates. In the last section, I briefly canvassed what I take to be the most plausible ways out of this puzzle, but more work must be done to concretise, evaluate, and combine these proposals so as to reinforce the foundations of the representational theory of mind. I hope that setting the puzzle and the possible solutions out as explicitly as possible will spur research by philosophers and cognitive scientists aimed at just this.[[17]](#footnote-17)

# References

Akins, Kathleen. (1996). Of sensory systems and the “aboutness” of mental states. *The Journal of Philosophy*, *93*(7), 337–372.

Akins, Kathleen A., & Hahn, Martin. (2014). More than mere colouring: The role of spectral information in human vision. *The British Journal for the Philosophy of Science*, *65*(1), 125–171.

Asoulin, Eran. (2020). Why should syntactic islands exist? *Mind & Language*.

Averill, Edward Wilson. (2005). Toward a projectivist account of color. *The Journal of philosophy*, *102*(5), 217–234.

Barwich, Ann-Sophie. (2019). A critique of olfactory objects. *Frontiers in psychology*, *10*, 1337.

Block, Ned. (1986). Advertisement for a semantics for psychology. *Midwest studies in philosophy*, *10*, 615–678.

Brandom, Robert. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard university press.

Burge, Tyler. (1986). Cartesian error and the objectivity of perception. In P. Pettit, & J. McDowell (Eds.) *Subject, Thought and Context*, (pp. 117–136). Oxford University Press Oxford.

Burge, Tyler. (2003). Psychology and the environment: Reply to chomsky. In M. Hahn, &

B. Ramberg (Eds.) *Reflections and Replies: Essays on the Philosophy of Tyler Burge*.

MIT press.

Burge, Tyler. (2005). Disjunctivism and perceptual psychology. *Philosophical Topics*, *33*(1), 1–78.

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

Byrne, Alex., & Hilbert, David R.. (2003). Color realism and color science. *Behavioral and Brain Sciences*, *26*(1), 3–21.

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

Chirimuuta, Mazviita. (2015). *Outside color: Perceptual science and the puzzle of color in philosophy*. MIT press.

Chomsky, Noam. (1959). A review of BF Skinner’s Verbal Behavior. *Language*, *35*(1), 26–58.

Chomsky, Noam. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Chomsky, N.. (2001). Derivation by phase. In M. Kenstowicz (Ed.) *Ken Hale: A life in language*, (pp. 1–52). Cambridge, MA: MIT Press.

Churchland, Patricia S., Ramachandran, Vilayanur S., & Sejnowski, Terrence J. (1994). A critique of pure vision. *Large-scale neuronal theories of the brain*, *23*.

Collins, John. (2008). A note on conventions and unvoiced syntax. *Croatian Journal of Philosophy*, *8*(23), 241–247.

Collins, John. (2014). Representations without representata: Content and illusion in linguistic theory. In P. Stalmaszczyk (Ed.) *Semantics and Beyond: Philosophical and Linguistic Inquiries*, (pp. 27–64). Walter de Gruyter.

Devitt, Michael. (2006). *Ignorance of language*. Oxford University Press on Demand.

Egan, Frances. (1995). Computation and content. *The Philosophical Review*, *104*(2),

181–203.

Egan, Frances. (2009). Is there a role for representational content in scientific psychology? *Stich and his critics*, *14*, 14.

Egan, Frances. (2014). How to think about mental content. *Philosophical Studies*, *170*(1), 115–135.

Egan, Frances. (2018). The nature and function of content in computational models. In *The Routledge handbook of the computational mind*, (pp. 247–258). Routledge.

Egan, M Frances. (1989). What’s wrong with the syntactic theory of mind. *Philosophy of Science*, *56*(4), 664–674.

Fodor, Jerry., & Lepore, Ernie. (2007). Brandom beleaguered. *Philosophy and Phenomenological Research*, *74*(3), 677–691.

Fodor, Jerry A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. New York: Random House.

Fodor, Jerry A.. (1986). Why paramecia don’t have mental representations. *Midwest Studies in Philosophy*, *10*(1), 3–23.

Fodor, Jerry A. (1987). The persistence of the attitudes. In *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT Press.

Fodor, Jerry A. (1990). *A theory of content and other essays.*. The MIT press.

Fodor, Jerry A. (2003). *Hume variations*. Clarendon Press.

Fodor, Jerry A., & Lepore, Ernest. (1992). *Holism: A shopper’s guide*. Wiley-Blackwell.

Fuld, Kenneth., Wooten, Billy R., & Whalen, James J. (1981). The elemental hues of short-wave and extraspectral lights. *Perception & psychophysics*, *29*(4), 317–322.

Gallistel, Charles R. (1998). Symbolic processes in the brain: The case of insect navigation. *An invitation to cognitive science*, *4*, 1–51.

Godfrey-Smith, Peter. (2001). Three kinds of adaptationism. In S. Orzack, & E. Sober

(Eds.) *Adaptationism and Optimality*. Cambridge University Press.

Greenberg, Mark., & Harman, Gilbert. (2008). Conceptual role semantics. In E. Lepore,

& B. Smith (Eds.) *The Oxford Handbook of Philosophy of Language*, (pp. 295–322). Oxford: Oxford University Press.

Hardin, CL. (2003). A spectral reflectance doth not a color make. *The Journal of Philosophy*, *100*(4), 191–202.

Hardin, Clyde L. (1988). *Color for philosophers: Unweaving the rainbow*. Hackett Publishing.

Hering, Ewald. (1878/1964). *Outlines of a Theory of the Light Sense -tr. Hurvich and Jameson*. Harvard University Press.

Johnson, G (this volume). *Unconscious Perception and Unconscious Bias: Parallel Debates About Unconscious Content*. In Kriegel, U. (ed.) *Oxford Studies in Philosophy of Mind vol. 3*. Oxford University Press.

Loar, Brian. (1981). *Meaning and mind*. Cambridge: Cambridge University Press.

Lycan, William G. (2014). The intentionality of smell. *Frontiers in Psychology*, *5*, 436.

Marr, David. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Mendelovici, Angela. (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical Studies*, *165*(2), 421–443.

Mendelovici, Angela A. (2018). *The phenomenal basis of intentionality*. Oxford University Press.

Millikan, Ruth Garrett. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT press.

Mitterer, Holger., & Blomert, Leo. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & psychophysics*, *65*(6), 956–969.

Newell, Allen. (1980). Physical symbol systems. *Cognitive science*, *4*(2), 135–183.

Pereplyotchik, David. (2017). *Psychosyntax: The nature of grammar and its place in the mind*, vol. 129. Springer.

Poldrack, Russell A. (2020). The physics of representation. *Synthese*, (pp. 1–19).

Pylyshyn, Zenon Walter. (1984). *Computation and cognition*. MIT press Cambridge, MA.

Quine, Willard Van. (1951). Ontology and ideology. *Philosophical Studies*, *2*(1), 11–15.

Rescorla, Michael. (2017). Levels of computational explanation. In *Philosophy and Computing*, (pp. 5–28). Springer.

Rey, Georges. (2006). Conventions, intuitions and linguistic inexistents: A reply to devitt. *Croatian Journal of Philosophy*, *6*(18), 549–569.

Rey, Georges. (2008). In defense of folieism: Replies to critics. *Croatian Journal of Philosophy*, *8*(23), 177–202.

Rey, Georges. (2020). *Representation of language: Philosophical issues in a Chomskyan linguistics*. Oxford University Press.

Schouten, Bert., Gerrits, Ellen., & Van Hessen, Arjan. (2003). The end of categorical perception as we know it. *Speech communication*, *41*(1), 71–80.

Shea, Nicholas. (2018). *Representation in cognitive science*. Oxford University Press.

Sprouse, Jon., & Hornstein, Norbert. (2013). *Experimental syntax and island effects*.

Cambridge University Press.

Stalnaker, Robert. (1989). On what’s in the head. *Philosophical Perspectives*, *3*, 287–316.

Sterelny, Kim. (2003). *Thought in a hostile world: The evolution of human cognition*. Blackwell Oxford.

Stich, Stephen P. (1983). *From folk psychology to cognitive science: The case against belief.*. the MIT press.

Witzel, Christoph. (2019). Misconceptions about colour categories. *Review of Philosophy and Psychology*, *10*(3), 499–540.

1. The claim is not that *all* representations must function in this way. Clearly, not all do. The hope is that an account of representation can be provided for some basic kinds of perceptual and cognitive processes, which can serve as a sort of “base case” from which other representational capacities and structures can be constructed. One key question here is how we get from the sorts of iconic representational capacities provided as paradigm cases here, for which structural relations in the representation (function to) correspond to structural relations in the represented target, to non-iconic representational systems, such as natural language. I will not take a stand on this issue in this paper, but will confine my discussion to cases that, at least *prima facie*, constitute this “lower level” of representational complexity. [↑](#footnote-ref-1)
2. What is controversial, and I deliberately leave open, is the status of this condition. Is it merely evidential, as Fodor (2003) suggests, or is the connection to different behavioural options constitutive of representation, as in Shea, following Millikan (1984) and Sterelny (2003)? [↑](#footnote-ref-2)
3. Burge’s focus on perceptual constancies stresses this aspect of representational psychology. [↑](#footnote-ref-3)
4. I do not assume anything too metaphysically robust here. I will count as sufficiently “lawlike” any connections between sensation and behaviour which enable scientific theories or models to explain the latter with reference to the former without appealing to distal features of the environment. What I will assume is that such a notion of “law” is intransitive, so that the existence of laws mapping sensation to representation (as posited in psychophysics) and mapping representations to behaviour (in line with the standard Fodorian view of psychological explanation), will not entail the existence of laws mapping sensation onto behaviour. [↑](#footnote-ref-4)
5. I say that positing representations is the *best* explanation of these capacities, not that it is the only possible explanation. Alternative proposals, such as *relationalism* (Campbell, 2002) instead explain such behavioural capacities with reference to the relations between the agent and the perceived distal object. This is not the place to enter into such debates, but I believe that the positing of *inaccurate* and *non-referring* representational states offers representationalism explanatory purchase not available to the relationalist. See Burge (2005) for extended discussion of this point. [↑](#footnote-ref-5)
6. This is of course massively simplified, both in the description of how texture gradients serve as cues for depth, and in abstracting from the way that texture along with many other cues are integrated. The point is merely to illustrate how paradigmatic representational explanation works. [↑](#footnote-ref-6)
7. That colour and speech are perceived categorically is about as close to consensus as one gets in psychology, but even here there are detractors. See e.g. Schouten et al. (2003) and Witzel (2019). [↑](#footnote-ref-7)
8. While the precise location of the unitary hues varies from subject to subject, that each individual takes some instance of red, yellow, green, and blue to be basic is widely accepted in the vision science literature. This observation dates back to Hering (1878/1964). See Hardin (1988) (pp. 39–45) for discussion and Fuld et al. (1981) for experimental results. [↑](#footnote-ref-8)
9. This prohibition can be overridden by adoption of words from other languages, such as ‘Tsunami’, ‘Tsar’, or ‘Tsetse’, but such words are easily recognisable as loan words and further the initial consonant is often dropped either as a matter of policy or in rapid speech. [↑](#footnote-ref-9)
10. Strictly, Japanese is typically viewed as based not around syllables, but around *moras*, a slightly different notion incorporating constraints on speech rhythm in addition to phoneme grouping. The difference shouldn’t matter for my purposes, and so I will stick to the more familiar notion of the syllable. [↑](#footnote-ref-10)
11. For my purposes, I can remain silent on the question of whether there is a more abstract way in which these stimuli can be said to have these properties. Devitt (2006), for example, argues that syntactic and semantic properties are genuinely properties of public linguistic objects, but they are “high level” relational properties held in virtue of a given public symbol’s relationship to the standing linguistic conventions of a linguistic community. Whatever the merits of this proposal, I can grant it here and still make the point that such properties are not plausibly causally responsible for the representational functioning of the speech perception systems, and so such properties seem ill-suited for featuring in the sort of correspondence-based views discussed above. [↑](#footnote-ref-11)
12. Byrne and Hilbert go on to precisify their view, and argue that reflectances are one ``flavor’’ (of colour), so as to allow for the existence of coloured light sources and coloured translucent volumes. They dub this larger category `productances’. This wrinkle will not matter for my purposes. [↑](#footnote-ref-12)
13. See also Johnson (this volume, section 3.1) for a clear discussion of the ways that colour vision science centrally involves appeal to perceptual *successes* and *failures*. [↑](#footnote-ref-13)
14. One prominent alternative is Mendelovici’s (2018) “phenomenal intentionality” approach, which aims to derive representational properties from properties of phenomenal experience. While discussion of this approach would take me too far afield for the purposes of this paper, it is worth noting that, especially in the case of speech perception, many of the purported representational properties (e.g. representation of a stimulus as an unpronounced subject, such as the subject of ‘dance’ in ‘Asher wants to dance’) are not consciously accessible, and so this strategy will not be available here. [↑](#footnote-ref-14)
15. One could of course endorse a pluralism about mental representation without agreeing with Egan that states in, say, the visual system were generally individuated without reference to the environment. Plausibly, states representing depth or object-boundaries might be externalistically individuated, while states representing colour and speech were individuated purely with reference to their computational roles. [↑](#footnote-ref-15)
16. Wh-islands are sentential contexts which preclude certain sorts of grammatical dependencies. For example, conjunctions generally preclude the extraction of a question-particle from only one of their conjuncts. Witness: \*’Who did Matthieu and \_\_\_ see at the bar?’. The core idea of Chomsky’s proposal is that computational efficiency is increased if our linguistic system identifies certain structures as complete, and thus do not allow any further operations on them. Without this constraint on interpretive dependencies, increases in sentence length will lead to explosions in computational complexity as each structural constituent could potentially depend on arbitrarily many others. Although see Sprouse & Hornstein (2013) and Asoulin (2020) for discussion and alternative accounts. [↑](#footnote-ref-16)
17. This paper benefited from feedback from Zed Adams, John Dupré, Gabbrielle Johnson, Uriah Kriegel, and Kevin Lande, as well as an audience at the British Society for the Philosophy of Science. [↑](#footnote-ref-17)