Holism, conceptual role, and conceptual similarity

Introduction

Many holistic views of mental content claim that content is not shareable: no two subjects ever mean the same things by what they say; nor can they entertain the same thoughts. This presents these views with some pressing concerns. There are certain phenomena that appear best explained by appeal to shared content. For example, one might think that successful communication requires that the hearer grasp the very content as that which is expressed by the speaker; and certain kinds of disagreement might require that two subjects take different attitudes towards the same content. These problems are serious: a theory of content that cannot underpin explanations of communication and disagreement is, at best, highly revisionary and, at worst, a non-starter. In response to these kinds of concern, holists often appeal to similarity of content. Similarity of content, it is thought, is an acceptable surrogate for sameness of content in explanations of the role of mental content in various intersubjective and intrasubjective activities and phenomena. Thus, providing an account of what it is for two contents (and the concepts they contain) to be similar is vital for many holists. Moreover, this account must be robust – that is, the account must define conceptual similarity for more than just the case of identical holistic networks. Fodor and Lepore (1992, 1999) have argued that this cannot be done. They argue that any account of what it is for two concepts (in non-identical networks) to be similar must presuppose a robust notion of conceptual identity. Holists, however, typically claim that there is no such thing as robust conceptual identity. If Fodor and Lepore are right then the holist is in trouble.

There are responses to Fodor and Lepore in the literature. However, existing accounts of how to compare concepts for similarity are only available to theories of concepts that are
partly, or fully, externalist.\footnote{An exception is Schroeder (2007), who provides a measure of conceptual similarity in terms of similarity of extension. His approach is compatible with both externalist views and the internalist view I defend in this paper. However, his approach compares concepts only indirectly – via their extensions. My aim is to provide an account of what it is for the conceptual roles of two concepts in different holistic networks to be similar.} For example, Paul Churchland’s State Space Semantics is a holistic theory; Churchland offers an account of conceptual similarity but his account, if successful, is only available to views that posit both an internal \textit{and} external determiner of content (Churchland 1998: 8).\footnote{Gärdenfors (2000) offers an account that is similar to Churchland’s, yet fully internalist. However, Gärdenfors does not offer a response to the present objection from Fodor and Lepore.} In this paper, my aim is to provide an account of robust conceptual similarity which is available to holists who endorse a purely internalist theory of mental content. My account appeals to similarities between perceptual representations in the cognitive systems of subjects, which can be used to ground comparisons of concepts across different holistic networks. In this paper, I focus on providing an account of conceptual similarity for holists who endorse internalist conceptual role theories of concepts in particular. However, I think a version of my account could be endorsed by other kinds of holist internalist as well.

The paper proceeds as follows. In Section 1, I introduce the kind of holistic theory that I will defend. In Section 2, I introduce Fodor and Lepore’s objection. In Sections 3 and 4, I present my solution.

\section{Holism, internalism, and conceptual role}

Conceptual role theories of concepts claim that concepts are individuated by their position in a network.\footnote{The accounts most similar to the one defended here are Rapaport (2002) and Hudson (2007). Other defenders of various kinds of conceptual role theory are Harman (1987) and Field (1977). Closely related are inferential role theories – see, e.g., Block (1986).} This position is determined by its relations to the positions of other entities in that network (where the network may include other kinds of entity in addition to concepts). All concepts are represented as nodes in some particular position in a network. These nodes are connected via a limited variety of links. Links between nodes represent special kinds of
relation, which serve to structure the network. These network-structuring relations, unlike concepts, are not holistically individuated by their location in the network. They are shared across networks, and are quite different from what we might call ‘relation-concepts’ (that is, concepts of relations, such as LOVES, RESEMBLES, KILLS, etc.), which, like any other kind of concept, are represented as nodes in the network, rather than as links between nodes. Theories as to the different kinds of network-structuring relation that concepts can bear to one another have been developed in linguistics (See, for example, Hudson 2007). One important kind of relation is that of category membership. For example, a subject’s CAT concept may be related to her ANIMAL and PET concepts, not merely associatively, but such that CAT is a member of the categories ANIMAL and PET; network-structuring relations can thus be directional (ANIMAL and PET are not members of the category CAT). Concepts can compose to form sentential contents. The relations between concepts (such as relations of category membership) determine how the sentential contents which they compose interact with each other in the cognitive system (e.g., which inferences they participate in) and how new information is used to update the network.

The particular version of conceptual role semantics that I will defend in this paper has two distinctive features. Firstly, it is fully holistic. Conceptual role theories can be more or less holistic depending on whether it is the total network, or some subset of it, that is responsible for individuating a particular concept. A fully holistic theory is one that claims that concepts are individuated by their relations to all other entities in the network. The second feature of the thesis defended in this paper is content internalism. This is the view that, for any subject, concepts are individuated solely by factors internal to that subject. I will

---

4 Network-structuring relations may be individuated in one of the ways considered in Section 4. That is, by their physical properties, or narrow functional role. It is worth noting here that, if one opts for the latter, one need not claim that these relation-types are holistically determined, even if a specification of their functional role includes their relations to concepts in the network. This is because the specific contents of concepts that are linked is irrelevant to a relation’s functional role. Network-structuring relations are neutral in this regard – their job is to link concepts of any type to other concepts of any type in a particular way – e.g., categorisation (cf. Schneider 2009: 538).
call the combination of these three theses – conceptual role theory, holism, and content internalism – ‘Holism’ for short. In the next section, I explain in more detail why the Holist needs an account of conceptual similarity and introduce Fodor and Lepore’s objection.

2 Holism and the problem of conceptual similarity

The Holist needs an account of conceptual similarity because, as Fodor and Lepore (1991) have argued, she must abandon the thesis that any two subjects with non-identical conceptual networks can possess type-identical concepts. That is, Holists must deny the following thesis:

Robust Content Identity: It is possible for a concept, C1, in a network, N1 to be type-identical to a concept, C2, in a non-identical network, N2.

Subjects who possess identical conceptual networks can share concepts, according to Holism. However, this is because network-identical subjects share all their concepts. The problems arise when we are comparing two subjects with non-identical conceptual networks. The reason for this is that Holism is thought to entail what is called the ‘Instability Thesis’:

Instability Thesis: Any change, however minute, to a subject’s conceptual network will determine a change in all concepts within that network.

Commitment to the Instability Thesis precludes Holism from endorsing Robust Content Identity because the Instability Thesis entails that, when two subjects possess different conceptual networks, they share none of their concepts. As mentioned above, this result is thought to pose serious problems for the Holist’s accounts of communication, disagreement, and translation, amongst other things, because explanations of these phenomena are often
thought to require that subjects share concepts (Fodor and Lepore 1991: 331). The Holist must explain these phenomena without appeal to shared concepts, or accept the unpalatable thesis that only subjects with identical networks can communicate, disagree, etc.\(^5\)

It is in response to this worry that Holists have proposed appealing to conceptual similarity. The Holist’s claim is that we can provide criteria that determine, for any two concepts in non-identical networks, the manner or degree to which they are similar. It is hoped that appeal to conceptual similarity can alleviate any costs incurred by Holism’s commitment to the Instability Thesis: similarity of content can play much the same role that sameness of content plays in traditional semantic theories vis-à-vis explaining communication, disagreement, translation, etc. (See, for example, Block 1986; Churchland 1986; Jorgensen 2009; Rapaport 2003). Thus, it is of the utmost importance to the Holist that she can provide an account of conceptual similarity that can figure in such explanations.\(^6\) In the remainder of this section, I explain why Fodor and Lepore (henceforth, ‘FL’) believe that no such account will be forthcoming.

FL (1992) argue that holists cannot provide criteria for robust conceptual similarity. This is now a long-standing problem for the Holist, and has been pressed by Fodor, sometimes with Lepore, in a number of places (e.g., Fodor 1998: Chapter 2; and FL 1999). Their argument is that any explication of conceptual similarity outside of identical networks must presuppose the Robust Content Identity thesis introduced above. Holism, as we have seen, is incompatible with this thesis. Thus, FL argue, even if the unsavoury consequences of the Instability Thesis could be alleviated by appeal to conceptual similarity, this option is not available to the Holist: we simply cannot make sense of conceptual similarity on the Holist’s view.

---

\(^5\) Some authors reject the charge that holism entails the Instability Thesis. See, for example, Pagin 2006 and Jackman 1999.

\(^6\) It also needs to be shown that conceptual similarity can play the required role in theories of communication, disagreement, etc. Demonstrating this is beyond the scope of this paper. However, Pollock (2015) argues that views of content that appeal only to conceptual similarity are actually better at explaining communication.
Their reasoning runs as follows. Suppose we ask, of a given concept, A1, in network, N1, whether it is similar to a second concept, A2, in a different network, N2. The Holist might answer by saying that A1 and A2 stand in similar conceptual relations to other concepts in their respective conceptual networks. But by what criteria are these conceptual relations similar? FL suggest that the only way to make sense of this idea is to cash it out as the claim that A1 and A2 are related to similar sets of concepts. That is, perhaps A1 is connected to concepts B, C, and D; and A2 is connected to concepts B, C, and E. But to claim this is to posit a robust notion of conceptual identity – we have posited tokens of type-identical concepts, B and C, across the two non-identical networks. FL claim that, in general, an explanation of similarity between two objects must appeal to overlap in the properties that these objects share. As such, FL think that there is no way to avoid positing some kind of robust conceptual identity in an attempt to provide criteria for conceptual similarity, and so the Holist must accept conceptual similarity as a primitive relation (FL 1992: 20). To be clear, it’s not just that providing such criteria would be hard – they think that it is not possible. Their claim is that “you can’t have a robust notion of content similarity … unless you have a correspondingly robust notion of content identity.” (1992: 199, emphasis in original) My central aim in this paper is to demonstrate how the Holist can get by without this appeal to conceptual identity.

3 A Response for Holism

Perhaps FL are right that we cannot give illuminating criteria for conceptual similarity by appealing only to the other concepts within a network. However, there is no reason that we should be forced to give our explanation purely in terms of concepts. Indeed, I think this would be a strange thing to attempt. The conceptual networks that underlie human natural language are not free-floating networks of arbitrary symbols. Rather, we should understand
them as networks that are embedded in, and intimately connected with, a complex broader system that includes other kinds of representation stored in the subject’s memory (cf. Hudson 2007: 18). In particular, conceptual networks are connected to an enormous number of perceptual representations, which represent things like the shape, colour, and texture of objects, the sounds that they make, the way they move, and so on. These representations are partially responsible for determining the conceptual role of the symbols in the cognitive system – how they interact with each other, how they respond to sensory inputs, and how they contribute to causing behavioural outputs.

In appealing to perceptual representations as distinct from concepts, my solution to FL’s challenge is committed to a distinction between conceptual and non-conceptual content – that is, in addition to positing concepts, the view posits a kind of representational content (enjoyed by perceptual representations) that is different in kind from the content of concepts. To track this in what follows, I will talk of both concepts and perceptual representations as inhabiting a ‘content network’ (where this should be understood as a network that stretches beyond a mere network of concepts). Like concepts, perceptual representations occupy unique positions in the content network. In addition to their connections to concepts, perceptual representations also have properties that are not determined holistically. That is, some of their properties do not depend holistically on their relations to other representations.

---

7 Whether there is such a thing as non-conceptual content is a topic of much debate, as is the question of how to mark the distinction between the conceptual and the non-conceptual. Defenders of non-conceptual content include Peacocke (1992), Bermudez (1995), Tye (2006), Raftopoulos and Müller 2006, and Jacob & Jeannerod (2003). Some authors are critical of the distinction (McDowell 1994, Stalnaker 1998, Dell’Anna & Frixione 2010). My use of the distinction has most in common with those who treat perceptual and subpersonal states as non-conceptual. It’s worth noting, however, that there are authors who would place perceptual representations on the conceptual side of the divide. Notably for my account, Barsalou (1999) and Prinz (2002) might be construed as taking at least some kinds of perceptual representation to be conceptual; my approach, in other respects, has much in common with these authors’ theories, but would be incompatible on this point. It should also be noted that, although Fodor (2008: 169ff) does maintain a distinction between conceptual and non-conceptual content, his division of representations might be at odds with the one I draw here. Fodor takes non-conceptual content to be iconic, where this amounts to the claim that, although a representation may have parts, it has no canonical decomposition. My account can be made compatible with his way of drawing the conceptual/non-conceptual distinction only if I can appeal to accounts of visual perceptual representations that treat them as iconic in this sense. As such, Fodor might have independent reason to reject my solution to FL’s challenge. Thank you to two anonymous reviewers for these points.
in the network. Thus, the conceptual roles of which they are a part also possess some non-holistically determined properties. It is certain of these non-holistically determined properties of perceptual representations that I will appeal to in my account of conceptual similarity. The basic idea, to be spelled out below, is that similar perceptual representations can be located and compared unproblematically across non-identical holistic networks, and these representations can be used to ground comparisons of holistically-determined concepts.

In Section 4, I will have more to say about the nature of perceptual representations. For now, I will work with the simplifying assumption that perceptual representations are often type-identical across holistic networks; we can think of ‘perceptual representation’ as a placeholder for some kind of entity which can play this role. Working with this assumption, I will demonstrate how this allows us to understand conceptual similarity across holistic networks. I will then argue that a slight complication of the assumption is both plausible and capable of playing the required role in my account of conceptual similarity. It will turn out that the holist can get by with the weaker claim that tokens of perceptual representations in different networks are not often type-identical, but can nonetheless be compared for similarity.

How can perceptual representations (henceforth ‘p-representations’) help us to ground conceptual similarity? There are different approaches one could take here and, importantly, there is one existing account in the literature that is easily confused with the one that I will develop. This existing account is a Humean empiricist view according to which concepts are built up out of p-representations. This Humean account has some things in common with my own approach, but it is not fully holistic. To avoid confusion, I will first briefly explain the Humean account and use this as a foil against which to highlight the structure of my proposed view. My hope is that setting things out this way will make clear precisely what is required to construct an account of robust conceptual similarity for a thoroughgoing Holist.
3.1 Humean empiricism

A version of the Humean approach is outlined and defended in Prinz (2006). The approach treats all concepts as ultimately derived from perceptual primitives. As Prinz puts it, the idea is to identify concepts with ‘collections of sensory features’ (2006: 101). At the simplest conceptual level, a concept might be derived directly from one or more perceptual primitives. These simple concepts may then themselves be used in the construction of more complex concepts, and so on in increasing levels of complexity. What this kind of theory has in common with the view I will develop is its empiricism: both views claim that many (and perhaps all) concepts are traceable to their roots in collections of perceptual representations.

It can also be described as a structured network view of sorts: on the Humean view, concepts are derived from entities that sit ‘below’ them in a network. The important difference between the two kinds of view, for my purposes, is that the Humean network is not fully holistic. The reason for this is that, although concepts depend for their individuation on the concepts and perceptual primitives that sit below them in the network structure, the set of representations relevant to the individuation of each concept is in most cases only a proper subset of the total set of representations in the network. A concept cannot, for example, be affected by changes to a further concept in the same network that is derived from a different set of perceptual primitives. Thus, although the view allows for a significant amount of interdependency amongst concepts, it does not entail the Instability Thesis. For a view to be

---

8 As Prinz notes, FL think that Churchland is covertly committed to something like this view. Although, it is worth noting that Churchland claims to be opposed to empiricism, as well as claiming to be a radical rather than moderate holist. See FL 1992: 189ff.

9 In addition to thinking that Holism cannot explain conceptual similarity, FL are also vehemently opposed to empiricism. There is not space in this paper to defend empiricism. However, it’s worth noting that my view is compatible with the claim that at least some non-concept entities in a network are native or non-perceptual. As such, I am not committed to the claim that all concepts must be rooted in nothing but perceptual representations. For a sustained defence of empiricism, see Prinz 2006 and 2002. As an anonymous reviewer points out, one thing that I am committed to is the claim that all concepts (if they are to be comparable with other concepts) must be connected to entities other than concepts (either directly or via intermediate concepts). That is, if there could be such a thing as isolated networks of pure concepts, my solution will not work for such networks.
fully holistic, it must be that a change to any concept in the network effects a change in all concepts in the network. Commitment to the Instability Thesis is a constraint on a successful solution for the radical Holist.

When it comes to explaining conceptual similarity, the Humean has a much easier time than the Holist. The Humean can appeal to overlapping sets of perceptual primitives to ground conceptual similarity between two concepts: two concepts are similar to the extent that there is overlap in the sets of representations from which they are constructed. Moreover, because of this, the Humean does have available to her a robust notion of content identity. Two networks may be non-identical in virtue of comprising different sets of perceptual primitives. However, these two networks may nonetheless share some proper subsets of perceptual primitives. If the same proper subsets of primitives are used to construct a concept in each of the networks, these concepts can be type-identical despite broader differences between their respective networks. ¹⁰ For the Holist, in contrast, all perceptual primitives are relevant to the individuation of all concepts in a network. We cannot carve out a proper subset that is relevant to one particular concept and not another. The Holist, then, will need to appeal to something other than overlapping sets of perceptual representations in explaining conceptual similarity. In the next section, I explain the approach that I think she should take.

### 3.2 Similarity across fully holistic networks

Consider the following, very simple, example. Suppose that in network N₁, we find p-representation S. Then suppose we find that some relationship holds between S and a concept token A₁. We can then look for a concept in a second network, N₂, which is related in the...
same way to a token of the same p-representation in this second network. These networks are represented in Fig. 1.

[Figure 1 here]

**Figure 1** Two holistic networks, N1 and N2. Square boxes represent concepts. Circular boxes represent p-representations. Arrows represent relations between elements of the network.

The fact that A1 and A2 are both connected to S tells us *something* about their similarity – it tells us that they are similar in this one respect. This is a good start. However, it does not tell us very much, for it may be the case that their relations to other concepts in the network render them dissimilar in other respects. For example, suppose we investigate the network further and find that connected to concepts B1 and B2 are two different p-representations: connected to B1 is p-representation P, and connected to B2 is p-representation Q. This is represented in Fig. 2.

[Figure 2 here]

**Figure 2** Two holistic networks, N1 and N2. Square boxes represent concepts. Circular boxes represent p-representations. Arrows represent relations between elements of the network.

Now we can say that concept A1 is similar to A2 in one respect, but differs from A2 in these two newly discovered respects. We now have a method for discerning when two concepts are similar or different *in one or more respects*. This is the core of my response to FL: we can say how many similarities and differences there are between two concepts, and we can specify the respects in which they are similar or different from each other. Using this approach, we can compare concepts across any number of networks. Consider Fig. 3.

[Figure 3 here]

**Figure 3** Three holistic networks, N1, N2 and N3. Square boxes represent concepts. Circular boxes represent p-representations. Arrows represent relations between elements of the network.
In Fig. 3, A1 (in N1) and A2 (in N2) are similar in two respects. Firstly, A1 and A2 are connected to p-representation S. Secondly, A1 and A2 are connected to concepts C1 and C2, respectively, which are themselves both connected to tokens of p-representation X. However, A1 and A2 also differ in two respects: A1 is connected to concept B1, which is connected to p-representation P, whereas A2 is connected to B2, which is connected to p-representation Q.

In contrast, concepts A1 (in N1) and A3 (in N3) are similar in only one respect: A1 and A3 are both connected to p-representation S but, other than this, they are dissimilar in four respects: A3 is connected to concept C3, which is connected to p-representation Y, and concept B3, which is connected to p-representation R; A1 is not connected to concepts with links to these p-representations; rather it is connected to two concepts with links to two different p-representations (P and X). Given this, we can say that A1 is more similar to A2 than it is to A3.

Note how this approach differs from the Humean approach introduced earlier. The Humean would use p-representations as a means of type-identifying concepts across networks. For example, the Humean view could treat concepts A1 and A2 as tokens of the same type in virtue of their relations to p-representation S. If this were the proposal, then we would be doing exactly what FL claimed we were forced to do: that is, we would have appealed to a robust notion of conceptual identity in our comparisons of concepts across networks. The present approach does not type-identify concepts across networks by appeal to some privileged class of representations: we are not shirking our commitment to the Instability Thesis. It will remain the case that any change to the network (either a change in concepts or in p-representations, or in link types) will determine a change in all concepts within the network. In Fig. 3, for example, concept A1, in N1, and concept A2, in N2 are not tokens of the same concept type. In accordance with the Instability Thesis, this is entailed by the fact that they inhabit networks that are not identical: the conceptual role of A1, in N1, is
affected by the presence of concepts B1 and C1; the conceptual role of A2, in N2, is affected by the presence of concepts B2 and C2. As such, although A1 and A2 do share some of their properties – for example, they are both connected to a token of p-representation S – they possess other properties which are not shared and so they enjoy different conceptual roles (occupy different locations) in the total network. As stressed above, although on this proposal the conceptual roles of concepts include elements which are not holistically determined, it does not follow that concepts are themselves not holistically determined. For a concept to be holistically individuated, it must simply be individuated by its connections to all other elements in the total network; and this is consistent with some elements of the network having non-holistically determined properties.

3.3 Varieties of link

In the previous examples, I appealed to different kinds of relation: relations between different concepts, as well as relations between concepts and p-representations. These were represented in the diagrams as different kinds of link between nodes. These are network-structuring relations in the sense introduced earlier. Precisely which network-structuring relations do these links represent? Let us start by considering the concept-to-concept relations. As we will see, these relations introduce complications that must be addressed. Consider the following networks in Fig. 4.

[Figure 4 here]

Figure 4 Two holistic networks, N4 and N5. Square boxes represent concepts. Circular boxes represent p-representations. Arrows represent relations between elements of the network.

In the above diagram, it might look tempting to say that A4 in N4, and A5 in N5 are similar in only one respect: their connection to p-representation S. However, I think consideration of the different kinds of relation that concepts can bear to one another gives us reason to think
that A4 and A5 have more in common than this. Recall that, earlier, I mentioned that one important kind of link in a conceptual network is one that represents the relation of category membership between concepts. Hudson calls this kind of link ‘Isa’; it is used in classification (for example, as in FODOR isa PHILOSOPHER, or CAT isa ANIMAL). On Hudson’s approach, conceptual networks are inheritance hierarchies built around isa relations connecting subcategories and supercategories. Subcategories inherit the properties of their supercategories. As Hudson illustrates, “For example, anything we know about Bird generalises to anything which isa Bird – in other words, to any particular bird or type of bird.” (Hudson 2007: 10–11) With this in mind, consider again the structure of the diagram from Fig. 4, but this time filled out with links that represent category membership. This is represented in Fig. 5 (I have labelled the concepts with more familiar names for illustrative purposes).

[Figure 5 here]

**Figure 5** Two holistic networks, N4 and N5. Square boxes represent concepts. Circular boxes represent p-representations. Arrows represent relations between elements of the network (black-headed arrows represent links between p-representations and concepts; white-headed arrows represent category membership between concepts).

Category membership is transitive: if CAT4 is a member of PET4, and PET4 is a member of ANIMAL4, then CAT4 is a member of ANIMAL4. Given this, I think the correct thing to say about the network represented in Fig. 5 is that there are actually two respects in which CAT4 and CAT5 are similar: they are both connected to p-representation S, and they are both connected (via a category membership link) to a concept which is connected to p-representation Q.

Hudson (2007) argues that there is more than one kind of network structuring relation. He posits five such relations in total. For our purposes, it is not necessary to go into the

---

11 It should be noted that there are well-established problems with viewing language networks in terms of inheritance hierarchies. Hudson deals with many such problems in his (2007).
details of Hudson’s particular version of network semantics. The purpose of mentioning his work is simply to illustrate the manner in which holistic networks can be structured by relations between concepts, and to demonstrate how appeal to such relations can help us to understand conceptual similarity across networks. Precisely which kinds of network-structuring relation structure the content network is an empirical matter, and one which I do not wish to take a stand on in this paper.\textsuperscript{12}

What about the relations between concepts and p-representations?\textsuperscript{13} Here there are different approaches one might wish to take. For example, one option would be to treat this relation as linking concepts with all stored p-representations of instances or exemplars to which the concept applies (we might call this the ‘instanted-by’ relation for short). For example, a subject’s DOG\textsubscript{S}\textsuperscript{14} concept might be linked, via this relation, to all p-representations taken to be representations of dogs, which have been stored in the subject’s memory following an encounter. Alternatively (or, perhaps, in addition), one may want to include links that select certain p-representations as default exemplars – those which are treated as more typical examples of the category.\textsuperscript{15}

The above was a highly simplified story of how we might go about comparing concepts for similarity across holistic networks. One thing to stress about the diagrams presented above is that the concepts represented by them are nothing like a typical human’s concepts. The diagrams represent extremely simple networks that, due to their size and the fact that they include few varieties of network-structuring relation, are significantly semantically impoverished. Even so, a human content network is just a more complicated

\textsuperscript{12}This is to say that it is an empirical matter which kinds of nodes and links best represent human content networks. For example, it might be that it is possible to construct a model that employs 5 kinds of link, but human content networks may be more accurately represented with 6 kinds of link. Thank you to an anonymous reviewer for suggesting this clarification.

\textsuperscript{13}Thank you to anonymous reviewer for suggesting clarification of this issue.

\textsuperscript{14}I use subscript here to indicate that this concept is idiosyncratic to the subject, S.

\textsuperscript{15}This is not to claim that concepts are exemplars. For examples of exemplar theories of concepts, see Medin and Schaffer (1978) and Estes (1994).
version of the networks represented in these diagrams. In practice, systematically comparing all the concepts in the content networks of two humans would be an arduous task. But this should not worry the Holist. FL claimed that there was some in principle difficulty in giving an account of robust similarity between concepts in holistic networks. They argued that we would not be able to do this without invoking robust conceptual identity. But it should now be clear that we do not need to do this. We can give an account of robust conceptual similarity in terms of similarities between concepts’ relations to the p-representations in the content network. We have a clear idea of how to go about comparing networks in the simpler cases. The human case is just a much more complex version of these simple cases.

In the preceding, I have explained how, on the assumption that p-representations are often type-identical across networks, we can compare concepts across holistic networks. However, as I mentioned when introducing the assumption, the Holist can get by with the weaker claim that p-representations, though not often type-identical, can be compared for similarity across networks. This weaker claim still needs support. To make this approach plausible, we need to do three further things. Firstly, we will need to explain what a perceptual representation is. Secondly, we will need to provide reason for thinking that such representations can themselves be compared for similarity across different networks. Lastly, we will need to demonstrate that this weaker claim is able to take the weight that is placed on it in my account of conceptual similarity. In the final section of this paper, I attempt each of these tasks.

4 Perceptual representations

4.1 Visual perceptual representations

Perceptual representations are representations stored in memory that encode perceptual information about objects, such as their shape, texture, orientation, sound, weight, bodily
location, etc. Perceptual representations have some properties that are not determined holistically and, as such, are not vulnerable to the destabilising effects of the Instability Thesis: some of their properties remain constant in the face of changes to concepts in the content network. It is commonplace for theories of concepts and word meaning to appeal to perceptual representations (see, for example, Prinz 2002, Barsalou 1999, and Jackendoff 1987). However, these representations have not yet been recruited for the purpose of explaining conceptual similarity in fully holistic networks – this is our present task. A complete account of holistic conceptual similarity ought to appeal to representations that record perceptual information across all sensory modalities. However, I will focus on vision as an example. Visual perceptual representations are records of the perceptual states that subjects undergo when visually perceiving a scene. However, they are not whole ‘snapshots’ of the perceived scenes. Rather, they record information about selected features of those scenes that can be later used in object recognition. One prominent view of how information is selected and stored following visual perception is Marr’s (1982) computational theory of vision. I will use Marr’s approach to provide an example of how perceptual representations could be used to ground similarity in holistic networks. My aim here is not to wed my approach to the success of Marr’s theory but, rather, to demonstrate how perceptual representations could, in principle, be usefully integrated into the Holist’s view.

Marr provides an account of how information from a two-dimensional array (such as the retinal image) can be used by a visual system to compute 3D models of the objects perceived. This account provides a description of the process of visual perception at the computational level, and is supposed to be applicable to any entity that is capable of visual

---

16 As mentioned above, Prinz (2006) does suggest something close to this approach, but his suggested view is not fully holistic – it allows that some concepts can be shared across non-identical networks. Furthermore, as also already noted, Prinz may object to my treatment of perceptual representations as non-conceptual.
perception.\textsuperscript{17} Marr (1982: 41ff) claims that this process involves three stages. The first is the primal sketch, which represents the presence and organisation of boundary elements from the two-dimensional retinal image. Information from the primal sketch, along with various assumptions built into the visual system, is then used to construct the 2½D sketch, which represents the geometry of surfaces – their depth and orientation – in a viewer-centred coordinate system (ibid: 268ff). Finally, this information (with the help of further assumptions) is in turn used to construct 3D models that provide a structural description of the spatial configuration of objects in an object-centred coordinate system (ibid: 295ff). These models represent the shape, volume, and orientation of objects in the visual field in three dimensions and, as such, represent features of these objects that are not all visible from the viewer’s perspective. For our purposes, what is important about Marr’s view is that the resultant object models are composed of more basic components. Object models are constructed out of a range of primitive elements organised in relation to a main axis. These primitives include generalised cones, which are linked together to represent parts of objects (1982: 223).\textsuperscript{18} For example, a human body could be represented by five generalised cones, suitably arranged, representing the head, torso, arms and legs. Each object perceived is decomposed into a hierarchy of these subcomponents, and the 3D object models can represent each level of the hierarchy down to an arbitrary level of detail (ibid: 302ff). This is represented in Fig. 6.

\textbf{Figure 6} A 3D model description of the human body, consisting of a hierarchically organised series of object models (Reproduced from Marr & Nishihara 1978, figure 3)

\textsuperscript{17} Marr (1982) also offers some suggestions concerning the algorithms employed in human visual perception and their neural implementation.

\textsuperscript{18} A generalized cone is the shape that is created by moving a cross section along an axis; the size of the cross section may vary along the axis, while its shape remains constant.
Perception and recognition of objects proceeds by constructing these structural descriptions of three-dimensional objects from the retinal image (via the stages described above) and then checking whether these representations match any of the representations that have been previously stored in the system.

We now have an account of what a visual perceptual representation is. Visual perceptual representations are hierarchically structured representations, which are produced as a result of visual perception and stored in memory. And, importantly, they include object models, which are produced by combining primitive elements. Now we are ready to turn to the second task of this section: we must make a case for the claim that such representations can be compared across subjects from within an internalist framework.

4.2 Comparing perceptual representations

How might perceptual representations be compared across subjects without reference to the external environment? Given the account of p-representations which is in play, I think the Holist should understand similarity of p-representations in terms of similarity between the combinations of primitive components that comprise them. The more similar the arrangement of components represented in two object models, for example, the more similar these object models are (cf. Prinz 2002: 170). At this stage, however, an objector may point out that, absent an account of the individuation of the primitive components, we have not offered a satisfactory solution to FL’s challenge. Indeed, it may look like we have just pushed the Holist’s problem to a different, equally problematic, location: haven’t we just replaced the problem of comparing concepts with the problem of comparing primitive perceptual representations? Whether this is so will depend on what account we give of the individuation of perceptual primitives.19 For example, suppose that these primitives are individuated, in

19 Thank you to an anonymous referee for pointing this out.
part, by their relations to concepts in the content network. If this were so, then we would indeed have simply relocated the original problem – for p-representations would not have any non-holistically identifiable properties that could be used to anchor the network in the way I am proposing. What we need is a way of individuating primitive representations such that they can be type-identified across individuals. It will be worth taking some time, then, to consider what options are (and are not) available to the Holist in this regard. In what follows, I will consider three options for typing representations across individuals. These are semantic accounts, physical accounts, and (narrow) functionalist accounts (See, for example, Aydede 2000, Pessin 1995, and Stich 1983).\footnote{My aim here is not to commit to a particular solution for the Holist, but to sketch the versions of these accounts that are available to her.} Let’s start with the semantic account. A semantic account types representations in accordance with their semantic interpretations. On this account, two token representations are of the same type if they have the same semantic content. There is ongoing debate concerning whether Marr’s theory is internalist or externalist. Burge (1986), for example, argues that the primitives involved in Marr’s primal sketch have externally individuated content. In contrast, Segal (1989) argues that the content of these primitives is internally determined. More recently, Prinz has endorsed an externalist semantic approach to the individuation of primitives, which treats them as individuated by the appearance properties in the world, which they detect (2002: 278). Obviously, the Holist (being an internalist) cannot appeal to externalist semantic individuation of perceptual primitives. If she wishes to maintain that primitives have semantic content, she must opt for the internalist interpretation. However, even assuming the truth of internalism about perceptual content,\footnote{There may be good reason to think that internalism about the content of perceptual primitives is false, of course. However, for present purposes it is reasonable to assume the truth of internalism and then see whether we have sufficient resources to explain conceptual similarity. If internalism about perceptual content is false, then perhaps Holism is false, but it is not the truth of internalism which is our concern in this paper.} the semantic approach is not of immediate help to the Holist when it comes to comparing primitives across individuals.
An objector will want to know which internal features are responsible for determining the content of primitives. And, once again, if these features include concepts in the content network then we are no closer to a solution to FL’s challenge than when we started. Fortunately, I think both the physical account and the functionalist account provide promising approaches to the individuation of primitives (although neither is wholly unproblematic). Either approach could be treated as the individuation base of narrow content for perceptual primitives; or, if one prefers, as non-semantic alternatives which may still play the same sort of role in the Holist’s account. I turn next to the physical account.

The physical account amounts to type-physicalism with respect to perceptual primitives: each type of perceptual primitive is identical with a type of physical (or quasi-physical) state. Two token primitives, across two different individuals, are of the same type if and only if they are of the same physical type, where a specification of this physical type may appeal to certain of the physical properties of neurons. Type-physicalism has been thought to be implausible for certain kinds of mental phenomena. For example, from the perspective of many theories, it will seem implausible as an account of specific concepts or sentential contents (Aydede 2000, Fodor 1994; although see Shagrir 1998: 450). One of the reasons for this is that many authors see concepts and thought contents as widely intersubjectively shared; but, so we might think, it is unlikely that the (shared) belief that DOGS BARK, for example, would be realised in the same type of physical state in two different subjects or, at least, in the same way across a large population: belief contents are multiply realizable. However, whether or not type-physicalism is plausible for concepts and thought contents, the present task for which the view is being recruited is different. We are interested

22 These non-semantic accounts can be characterised as different forms of syntactic individuation. I have avoided this terminology in what follows as authors relevant to this debate use the term ‘syntactic’ in different ways. For discussion of this issue, see Aydede (2000, 1997).
23 This terminology of from Aydede (2000); he points out that the relevant physical properties may be higher-order physical properties, such as physical shape.
24 Similar worries apply at the level of syntax - see Stich (1983).
in explaining how perceptual primitives can be shared across individuals. This is a more modest goal: whilst a human’s conceptual repertoire is vast, Marr lists fewer than 20 types of primitive representation in detailing his computational theory of vision (Marr 1982: 37). Appealing to type-physicalism in relation to these primitives involves claiming that types of perceptual primitive (blobs, boundaries, generalized cones, and so forth) are identical with particular types of physical state.\(^{25}\) This is certainly not uncontroversial – opponents of type-physicalism may insist that perceptual primitives are multiply realizable. The Holist who opts for a physical account must deny this; or, at least, accept that her account of conceptual similarity is restricted in its application only to comparisons between individuals who do indeed share the relevant physical characteristics. I think there is reason for the Holist to be optimistic. Many authors have forcefully challenged whether multiple realizability holds with respect to various kinds of mental state, as well as whether it poses a problem for type-physicalism (See, e.g., Shagrir 1998, Polger 2009, Shapiro 2000, Bickle 1998, Zangwill 1992, Bechtel and Mundale 1999, Couch 2004, Miłkowski 2016).\(^{26}\) Nonetheless, denying the multiple realizability of perceptual primitives would be a significant commitment, and one which some authors will see as too high a price.\(^{27}\) It would be nice if a type-physicalism was not the Holist’s only option. Fortunately, I think the Holist may also plausibly endorse a functionalist approach and, thus, it may be open to her to maintain that perceptual primitives are multiply realizable.

On the functionalist approach, a primitive would be individuated by its functional role in a particular system. Given her commitment to internalism, the Holist must here appeal to

\(^{25}\) This is not to say that a theory of vision should be ultimately reducible to the neurophysical level – Marr would have objected to this (Marr 1982: 18-19).

\(^{26}\) There are, of course, authors who defend multiple realizability for various kinds of mental state. See, for example, Putnam (1967), Aizawa and Gillett (2009).

\(^{27}\) It is also not the only problem with a physical approach to the individuation of representations. Pessin (1995), for example, offers further criticisms of the physical approach as it applies to symbols in the Language of Thought; it is not clear that these same criticisms would apply to the individuation of perceptual primitives, however.
narrow functional role, where this may include a representation’s relations to things like proximal stimuli, behaviour, and other representations. As we have noted above, for this to work in the Holist’s account of perceptual primitives, the individuation base of a primitive must not include holistically-determined concepts. There is reason to be worried here. Functional and computational role accounts of individuation for various mental phenomena are often thought (by both their advocates and their critics) to be holistic. For example, Schneider (2009) argues for holistic individuation of symbols in the Language of Thought – her view is that such symbols are individuated by their computational role in the central systems. Moreover, Fodor and Lepore (1991) object to conceptual role semantics, in part, because they see no principled way to prevent it from descending into radical holism. They argue that there is no non-arbitrary way to demarcate parts of the total network that are relevant to the individuation of a concept from those that are not without invoking an analytic/synthetic distinction – an option they see as untenable.\footnote{As Block (1993) suggests, even appeal to an analytic/synthetic distinction will not obviously help to determine unique conceptual roles for concepts.} However, fortunately, it is not clear that these same sorts of worries would carry over to the functional individuation of perceptual primitives. Perceptual primitives are individuated somewhat differently from concepts or symbols in the Language of Thought. Concepts (on Holism) are individuated by their locations in a network. If one wanted to avoid radical holism for concepts, one would have to find a principled way to demarcate some concept-individuating part of the network from the total network (for each concept). In contrast, whilst perceptual primitives may have a total functional role in the cognitive system, it does not seem arbitrary or unprincipled to select the visual system (or, perhaps, some subset of processes in the visual system) as the individuation base for (visual) perceptual primitives. The primitive representations in Marr’s theory, for example, have well-defined roles in the processes that produce 3D model descriptions from the retinal image. For example, primitives in the primal sketch (such as
bars, blobs, and terminations) are produced from combining oriented zero-crossing segments, whose job it is, in turn, to track intensity changes in the retinal image (Marr 1982: 404ff, Marr & Hildreth 1980).

In sketching these approaches to the individuation of perceptual primitives, I hope to have demonstrated that the Holist does have options here. The only options unavailable to her are (a) holistic individuation of perceptual primitives and (b) externalist semantic individuation. Before moving on, it is worth noting that, although neither of the Holist’s available options is unproblematic, many authors (including non-Holists) ought to have an interest in seeing one of her options succeed. As Schneider (2009) points out, those who favour a computational theory of mind need an account of the type individuation of symbols such that we can understand the claim that thought is a computational process that manipulates them. Schneider is primarily interested in the typing of symbols in the Language of Thought, but the same issue confronts perceptual primitives, considered as representations subject to manipulation in computational processes. It does not help to type symbols by their broad content for this task (even for those who posit broad content for perceptual primitives); this is because many authors think that broad content is not relevant to explaining the role of a symbol in computational processes (Schneider 2009: 526). One can still appeal to (narrow) holistic functional/computational role to individuate primitives (as Schneider does for symbols in the Language of Thought); but any author who wishes to maintain intersubjective generalizations about the role of specific types of primitive representation in visual perception will want to avoid this option, just as the Holist does.

In the preceding, I suggested that primitive perceptual representations can be type-identified across individuals – either by their physical properties or functional roles. We can understand similarity of complex representations, such as 3D model descriptions, in terms of similarity between these primitive components and their organisation. However, as various
authors have stressed, the complex perceptual representations that combine primitives are not likely to be often type-identical (See, for example, Prinz 2002: 153 and Barsalou 1999: 588). Individuals will have had more or less different perceptual experiences of the objects in the world, and thus will have stored representations that differ in both number and composition. We have now arrived at the final task of this section: in what follows, I will explain how appeal to mere similarity between (complex) perceptual representations can take the place of type-identical (complex) p-representations in our account of conceptual similarity.

4.3 Perceptual representations and conceptual similarity

I have just suggested that, although it is plausible that perceptual primitives are shared across individuals, the complex representations constructed from combining these primitives will rarely be shared. Because of this, we can no longer (or rarely) claim that any two concepts, across individuals, are connected to the same p-representations, as was assumed in presenting the account in Section 3. However, we can still make sense of the claim that these concepts are more or less similar in various respects. Whereas previously we described concepts as being the same or different in one or more respects (depending on their relations to type-identical p-representations), we should now say instead that two concepts are ‘highly similar’ or ‘moderately similar’ or ‘highly dissimilar’, etc., along certain dimensions. Two concepts are highly similar along some dimension when they are each connected to one or more p-representations which are themselves highly similar to each other. Two concepts will be highly dissimilar along some dimension when they are connected to p-representations that are highly dissimilar. We can say that a concept A1, in network N1, is more similar to A2 (in N2) than to A3 (in A3) just when (or to the degree that) there are higher similarities between A1’s connections to the concepts in N1 and A2’s connections to the concepts in N2 than there are between A1’s connections to the concepts in N1 and A3’s connections to the concepts in N3.
Replacing type-identity of (complex) p-representations with mere similarity (to some specified degree) between p-representations, then, should not cause problems for the Holist’s account of conceptual similarity. Indeed, it doesn’t matter if two subjects’ p-representations are highly dissimilar across the board: we will still have the resources to explain both how dissimilar they are, and in what respects.

Thus we have a response to Fodor and Lepore’s challenge. Fodor and Lepore objected that the Holist could not provide a robust account of conceptual similarity because, whatever account she gives must appeal to a correspondingly robust notion of conceptual identity, and the latter is not consistent with Holism. On my proposal, we can compare concepts by observing their relations to all of the p-representations in their respective networks. Crucially, these p-representations are such that they can be compared for similarity across networks in terms of their basic components. Two concepts, A1 and A2, across non-identical networks N1 and N2, are similar to the degree that they are connected (via the same network-structuring relations) to similar concepts and perceptual representations, where this degree of similarity should be measured relative to all concepts and perceptual representations in N1 and N2, respectively.

**Conclusion**

The purpose of this paper was to defend Holism from an objection from Fodor and Lepore. Holism is a conceptual role theory of mental content which is both fully holistic and purely internalist. FL’s objection was that the Holist cannot provide an account of robust similarity between concepts without presupposing a robust notion of conceptual identity. In response to the objection, I argued that we can provide an account of conceptual similarity by appealing to elements of content networks which have non-holistically determined properties; the elements I appealed to were perceptual representations. Concepts can be located and
compared by observing the relations that they bear to these perceptual representations – some direct and some via further concepts. As such, the Holist can appeal to a non-mysterious notion of robust conceptual similarity in her attempts to meet concerns pertaining to holistic accounts of communication, disagreement, and any other phenomena that are thought to be threatened by the Instability Thesis.

**Bibliography**


