

Non-compositional Representation in Connectionist Networks

Ronald L. Chrisley

School of Cognitive and Computing Sciences, University of Sussex
Brighton, United Kingdom

Abstract

I show that despite some recent arguments to the contrary, connectionist representations can often be non-compositional. This is not because they have context-sensitive constituents, but rather because they sometimes have no constituents at all. This is a consequence of a holistic approach to the ascription of representational content, which starts with propositional contents and works down, instead of assigning atomic sub-propositional contents and working up. But in order for such contents to be carried by unstructured representations, they must be non-conceptual.

1 Introduction

Conventional wisdom amongst connectionists is that human cognition is (at least sometimes) sensitive to context, and that it is so to such an extent that traditional models of cognition, which use the compositional, context-insensitive symbol as their explanatory workhorse, are doomed to failure, or at least to incompleteness. It is no surprise, then, that conventional wisdom amongst connectionists also has it that connectionist representations are (at least sometimes) context-sensitive and (thus) non-compositional, placing them in an excellent position to fill the explanatory void in traditional cognitive science that connectionist proselytizing has highlighted.

Unfortunately for connectionists, conventional wisdom is the stalking horse of the philosopher. In his paper “Content, Context and Compositionality”, Keith Butler argues 1) that empirical data typically cited by connectionists (e.g., Goschke and Koppelberg [13]) do not show that human mental representations are context-sensitive; and 2) that connectionist representations, be they microfeatural or recurrent, are, contrary to conventional wisdom, compositional.

Why should connectionists care about Butler’s claims? One reason is this: non-compositionality is the best way out of the notorious dilemma with which Fodor and Pylyshyn confronted connectionists: either connectionist nets are not systematic, and therefore cannot explain systematic cognition, or they are systematic, but only by being mere implementations of classical cognitive architecture, thus offering nothing new on the level of cognitive theory [12]. The popular connectionist answer to this dilemma has been to claim that connectionist networks can be systematic, but are not mere implementations of traditional cognitive architecture because their representations are context-sensitive

in a way that violates compositionality. So for most connectionists, to give in to Butler would be to give in to Fodor and Pylyshyn – an unattractive prospect.

Butler seems right on target concerning his first claim. He convincingly demonstrates that the linguistic data that Goschke and Koppleberg [13] cite, even if they are evidence for the non-compositionality of *linguistic expressions* (a claim which he also disputes), are not ipso facto evidence for the non-compositionality of *mental representations*. Nevertheless, Butler overstates his case when he concludes:

The evidence of context effects in linguistic representation implies no claims concerning context-sensitivity in mental representation. Consequently, we have found no warrant to challenge the semantic compositionality of mental representations. [2, p 22-23]

The flagship argument *for* the compositionality of mental representations relies primarily on the productivity and systematicity of language [12]. If it turns out that language expressions are context sensitive, and thus non-compositional, then it is open to the connectionist to pose a dilemma for the classical, compositional view: either compositionality is not required for systematicity and productivity in general, since it is not required for linguistic systematicity and productivity; or language is not truly systematic and productive. Either way, the demand for compositional mental representations is removed.

One could try to rebut Butler’s second argument by claiming that connectionist representational constituents are context-sensitive; are not compositional. However, it seems that what Butler means by “constituent” is so strong and particular that he is probably right in this respect: constituents, as he understands them, *must* be compositional, by definition/stipulation. So to avoid trivializing the debate, I will argue against Butler by showing that some propositional connectionist representations have no constituents at all. I will use Butler’s strong notion of constituent against him here: if something is only a constituent if it guarantees compositionality, then many connectionist propositional representations have no constituents. They therefore cannot be compositional.¹

2 No constituents, no compositionality

First, let’s be clear about what we’re talking about. Butler gives an idea of what he means by compositionality when he states:

Compositionality... requires that the meanings of complex mental representations are determined by the meanings of their component concepts, and the structural relationships between them. [2, p 4]

¹Note that I am not claiming that *all* connectionist representations are non-compositional, just the *some* are; but even this is enough to defeat not only Butler’s arguments, but his conclusions.

Despite the unconventional use of the word “concept”, and philosophical niceties aside, I’ll take this passage as confirming that Butler is subscribing to the standard notion of compositionality: that the meanings of complex representations are determined by the meaning of their parts.

It is clear from the above that compositionality requires constituent structure, as Butler admits:

Compositionality requires constituents of some sort, constituents that make uniform semantic contributions to the complex representations in which they figure. This is why Fodor and Pylyshyn (1988) were so concerned to establish the compositionality of mental representations; if there are no constituents, there is no compositionality. [2, p 18].

Thus, when Clark argues [6, pp 26-33] that the representations of simple recurrent networks (SRNs), such as Elman networks [11], have no constituents, this is a threat to Butler’s claim that all connectionist representations are compositional. Therefore Butler sets out to show first that such representations do have constituents, and then later that they are compositional. Specifically he considers the case of an SRN sequentially encoding an English sentence:

Take the familiar ‘John loves Mary’, fed to the network as a sequence of three patterns of activation across the input units. How are we to interpret the hidden-unit activation patterns generated by this input sequence? It would seem that we have two interpretations open to us: on the first, the complex propositional content attaches only to the *synchronic* hidden-unit activation pattern generated by the sentence-terminal word (Mary), on the second, the complex propositional content attaches to the *diachronic* sequence of hidden-unit activation patterns. [2, p 20, emphasis in the original]

Butler agrees with Clark that under the synchronic interpretation, compositionality can be denied on the basis that (at least some kinds of) connectionist representations have no constituents at all, let alone compositional constituents. Thus, Butler must argue against the synchronic interpretation of SRN representations.

Interestingly, Butler offers two reasons for *accepting* that it is the synchronic hidden unit activation pattern that is a bearer of representational content in an SRN: 1) it fits well with the synchronic aspects of our “working theory of mind”; and 2) “only synchronic aspects of the representation can be efficacious” [2, p 20].

Nevertheless, Butler claims that the synchronic interpretation should be rejected because it cannot explain how a representation has the content it does. The final hidden-unit pattern after encoding ‘John loves Mary’ can only be taken to mean that ‘John loves Mary’ by appealing to the fact that it is an encoding of a series of activation patterns, the first of which means ‘John’, the second of which means ‘loves’, and the last of which means ‘Mary’. That is,

only the diachronic interpretation can explain how a hidden-unit pattern has its content.

3 Holistic ascriptions of content

However, Butler does not seem to consider the possibility that a representation might have a propositional content directly, rather than derivatively by virtue of compositionality. That is, he begs the question in favour of compositionality, because he doesn't allow for any other possibility:

But while there may be no synchronic constituents that can contribute their contents to the complex propositional content, there are diachronic constituents... [a]nd these *do* contribute their contents to the complex content of the resulting activation pattern. Where else would the content of the occurrent hidden-unit pattern come from? [2, p 19, emphasis in the original].

Butler is assuming an *atomistic* theory of representational content, in which the contents of sub-propositional representations are fixed in advance (it is not clear how), and propositional contents are had by virtue of combining these syntactic atoms together in particular ways. But this is not the only theory of representational content. In particular, connectionist representations lend themselves well to being understood in terms of a *holistic* theory of representational content [4]. A holistic theory instead begins with the ascription of whole, propositional contents to the system in question (perhaps in an interpretational manner, such as Dennett's intentional stance [10]). A representational state (or other non-semantic aspect of the system) is then found which can serve as the vehicle of the propositional content. For example, it may be that, *ceteris paribus*, the possession of the content in question, and only that content, is counter-factually covariant with being in a certain representational state. It *might* then be possible to further decompose that representational state into constituents which counter-factually covary with the semantic constituents of the propositional content in question. But most importantly, *it might not*. A holistic theory leaves open the possibility that a propositional content may be carried by a representation that has no constituents.

The idea that a content can be carried by a representation, but not by virtue of any structure of that representation, might seem strange at first. But on further reflection, one realizes that even an atomistic theory must appeal to this idea. The contents that are carried by representational atoms cannot, by definition, be done so by virtue of the internal structure of those atoms. So if sub-propositional contents can be assigned directly to representations, why not propositional contents?

A natural response to this on Butler's behalf would be to point out that we were originally concerned with *conceptually structured* propositional contents, such as "John loves Mary". A system cannot be credited with the possession of

such structured contents unless it also has the capacity to possess related contents (such as those involving the concepts **John**, **Loves** and **Mary**). But the kind of system that I offered above as being non-compositional would not have these attendant capacities. For it could do so only if the representation carrying the propositional content ‘John loves Mary’ had constituents corresponding to the constituents of the content. Thus, I have not provided a counter-example, of a connectionist representation which has that conceptually structured propositional content and yet no constituents.

I have two replies to this. First, to make this point is to assume that the only way to exhibit systematicity is through compositionality of a very traditional sort. Yet this is a very contentious issue, and there is good reason to believe that this assumption is incorrect. It seems possible that the pressures of evolution, development and learning might ensure that a system is capable of behaving in such a way that it warrants the ascription of a capacity to possess the entire family of related propositions that conceptual structure demands, and yet the causal mechanisms which underly these capacities are distinct and unrelated. On a holistic theory of content, two semantically related contents can be carried by representations which have no common causal elements. Thus, we can make sense of a connectionist system which represents a conceptually structured propositional content in a non-compositional way.

The second reply is more concessive, but perhaps closer to the truth. Suppose that conceptually structured contents *do* require a mirroring representational structure, and thus compositionality. But then it seems that the kind of propositional content being ascribed in the holist case at hand must be one that is *not* conceptually structured. It must be what has been called *non-conceptual* content [1, 3, 4, 5, 7, 8, 9, 14] Connectionist representations with non-conceptual content, then, will typically be non-compositional. So Butler has only shown that connectionist representations of conceptual content must be compositional. But then the open question is: can cognitive science get by on conceptual content alone? If not (and there is good reason to believe that non-conceptual content *is* necessary for a proper understanding of cognition [8, 9]), then non-compositional connectionist representations are very much on the cards. They certainly have not been shown to be impossible.

4 Conclusion

Although Butler was mistaken on some points, he was correct to press us for an explanation of connectionist non-compositionality. For now we can see that if one wants to use non-compositionality as a means of getting between the horns of implementation and inadequacy in Fodor and Pylyshyn’s dilemma, then one will have to embrace both a holistic approach to the ascription of representational content, as well as a notion of non-conceptual content. With these in place, one can then gain a better understanding of how representations without constituents can nevertheless carry propositional content.

References

- [1] J. L. Bermudez. Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language*, 10(4), 1995.
- [2] K. Butler. Content, context, and compositionality. *Mind and Language*, 10(1/2):3–24, 1995.
- [3] R. Chrisley. Cognitive map construction and use: A parallel distributed processing approach. In D. Touretzky, J. Elman, T. Sejnowski, and G. Hinton, editors, *Connectionist Models: The Proceedings of the 1990 Connectionist Models Summer School*. Morgan Kaufmann, San Mateo, 1990.
- [4] R. Chrisley. Connectionism, cognitive maps & the development of objectivity. *Artificial intelligence review*, 7:329–354, 1993.
- [5] R. Chrisley and A. Holland. Connectionist synthetic epistemology: Requirements for the development of objectivity. In L. Niklasson and M. Bodén, editors, *Current Trends in Connectionism: Proceedings of the 1995 Swedish Conference on Connectionism*, pages 283–309, Hillsdale, NJ, 1995. Lawrence Erlbaum.
- [6] A. Clark. *Associative Engines*. MIT Press, Cambridge, 1993.
- [7] T. Crane. The non-conceptual content of experience. In T. Crane, editor, *The Contents of Experience*. Cambridge University Press, Cambridge, 1992.
- [8] A. Cussins. The connectionist construction of concepts. In M. Boden, editor, *The Philosophy of Artificial Intelligence*, pages 368–440. Oxford University Press, Oxford, 1990.
- [9] A. Cussins. The limitations of pluralism. In D. Charles and K. Lennon, editors, *Reduction, Explanation and Realism*, pages 179–223. Clarendon Press, Oxford, 1992.
- [10] D. Dennett. *The Intentional Stance*. MIT Press, Cambridge, 1987.
- [11] J. Elman. Finding structure in time. *Cognitive Science*, 14:179–212, 1990.
- [12] J. Fodor and Z. Pylyshyn. Connectionism and cognitive architecture: A critical analysis. In S. Pinker and J. Mehler, editors, *Connections and Symbols*. MIT Press, Cambridge, 1988.
- [13] T. Goschke and D. Koppelberg. The concept of representation and the representation of concepts in connectionist models. In W. Ramsey, S. Stich, and D. Rumelhart, editors, *Philosophy and Connectionist Theory*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.
- [14] C. Peacocke. *A Study of Concepts*. MIT Press, Cambridge, 1993.