Is Compositionality a Shortcut to Naturalization?

Derek Lam

November 8, 2017

Abstract

Philosophers have tried to naturalize intentionality by reducing mental representations in terms of certain non-intentional tracking relation. Various non-intentional tracking relations have been tipped for the job. Typically, it is considered enough that these non-intentional relations give us an analysis for simple mental representations because all other mental representations are composed of the simple ones. I examine two natural ways to interpret this claim and argue that, interpreted in either way, the compositional nature of mental representations is not a shortcut to naturalization.

1 Two-Step Naturalization

Intentionality is one of the two major obstacles to reductive physicalism about the mind.[1] The reductionist dream is to naturalize intentionality by finding a proper non-intentional feature that matches the theoretical profile of aboutness so that we can reduce the aboutness of our mind in terms of that feature. By doing so, we can leave intentionality out of reality. Fodor's much cited claim sums up the spirit of such a reductive approach: "If aboutness is real, it must really be something else." (1987: 97)[2]

Suppose our minds process non-semantically individuated representations. Suppose I have a simple mental representation type R. In virtue of what could R be *about cows*? Different stories have been told. For example, Dretske (1981), Millikan (1984), and Fodor (1990) represent three major approaches. Roughly put, Dretske's account says, for R to be about cows, it must be the case that, during the learning period of R, R is (typically) caused by cows. After the learning period, an instance of R is about cow regardless of its etiology. Millikan's teleosemantic account says that R is about cows just in case the evolutionarily intended situation in

---

[1]The other is phenomenal consciousness.

[2]Note that one does not have to be a physicalist to be a reductionist about intentionality. All it takes is to think that intentionality can be reduced to non-intentional features, which may or may not be physical.

which the subsystem that uses/consumes R is a situation in which R correlates with the existence of cows.[3]Finally, for Fodor, R is about cows just in case there is an asymmetric lawful correlation between R and cows, which means: R nomologically correlates with cows *and*, whereas some other mental representations can correlate with cows in virtue of R's correlation with cows, R's correlation with cows is not in virtue of some other mental representation correlating with cows. Schematically put, each of these naturalistic theories attempts to find a non-intentional relation T such that R is about cows if and only if T(R, cows).

For the sake of argument, let's suppose that these accounts work for simple cases like the concept ⟨cow⟩. Nonetheless, it's implausible that they work for our mental representations about many other things, e.g., the Big Bang, numbers, dinosaurs, etc. We don't seem to stand in any naturalistic relation to the Big Bang (which is too remote from us), numbers (which are, arguably, abstract), and dinosaurs (which do not exist anymore) — at least not in any naturalistic relation robust enough to account for the aboutness of mental representations.

Reductionists generally believe that such concerns can be swept aside by appealing to the compositional nature of these trickier mental representations. E.g., Ryder writes about representation of non-existing things:

> How can a cell refer to something that does not exist? To answer this sort of question, I would normally say that such representations are composed, and my neurosemantics is only meant to apply directly to atomic representations. (2004: 235; my emphasis)

Apparently, one only needs to naturalize the simple mental representations from which the rest are composed of. Dretske even considers the move evident enough that it doesn't need a defense:

> A primitive concept requires the system possessing it to have the capacity for receiving [...] information corresponding to the meaning of that concept (its semantic content). This, of course, is not true for complex concepts. [...] They are built up out of more elementary cognitive structures. Since I am not going to argue for this claim, I hope it is sufficiently plausible not to need argument. (1981: 222-223)

The goal of this essay is to critically examine this convenient reductionist response.

---

[3]This is surely a simplification of Millikan's complicated view. She believes that certain mental representations of complete states-of-affairs are primary and mental representations of objects and properties are derivative (Papineau (1987) holds a similar proposition-first view). I gloss over this aspect of Millikan's view because I believe the problem I am going to present for naturalization applies regardless of this complication. As long as a project to naturalize mental representation *depends* on the idea that some complex mental representations are results of (re-)combining simpler mental representations as components in some systematic way, what I say in this essay applies.

I suspect part of the temptation to think that the two-step approach works trivially — naturalizing the simple ones and then let compositionality do the rest — stems from conflating "mental representation" as a phrase referring to an entity that is a vehicle of content and "mental representation" as a phrase referring to the fact that such an entity has this or that content. The difficulty in reducing mental representation pertains to the reduction of the representational facts, not the things that carry representational content. To put the point in terms of linguistic representations, the reductionists' task is about naturalistically analyzing intentional facts like that "cows" is about cows, not naturalistically analyzing the word "cow". It's relatively uncontroversial that, if "milkshake" is an entity formed by sticking "milk" and "shake" together, which are non-intentional entities (as ink marks on papers), the composite "milkshake" is also a non-intentional entity. It's a different and far less straight forward question whether, as long as the fact that "milk" is about milk and the fact that "shake" is about the shaking motion are non-intentionally analyzable facts, the fact that "milkshake" is about milkshake is also non-intentionally analyzable.

The reductionist response has been taken for granted for a long time and it has never been made clear how exactly the compositional nature of mental representation helps us in the tricky cases. Worse, it isn't even clear how we're supposed to understand the claim that those tricky mental representations are composed.

The claim could be about what I'll call the syntactic compositionality of complex mental representations; namely, the idea that complex mental representations are constructed by piecing simpler mental representations together syntactically (i.e., by certain non-semantic relations). The claim could also be understood to be about the semantic compositionality of complex mental representations (which is what people usually mean when they speak of compositionality). It's the idea that the content of a complex mental representation is jointly fixed by the contents of the mental representations that are its components.

I'll argue that neither notions of compositionality can obviously help the reductionists. Hence, the matter isn't trivial at all.

2 Invoking Semantic Compositionality

Semantic compositionality says: for any complex mental representation R that is composed of the simpler mental representations $\{r_1, r_2, r_3,\dots\}$, what R represents is solely fixed by what the members of $\{r_1, r_2, r_3,\dots\}$ represent. Contents of the components jointly determine the content of the complex.

Suppose I have three mental representations R_1, R_2, and R_3; R_1 and R_2 jointly compose R_3. Suppose the representations are about the objects O_1, O_2, and O_3 respectively. Our reductionist offers a theory that says the fact that R_1 is about O_1 *just is* the fact that R_1 and O_1 stand in certain non-intentional tracking relation T. The same is true for R_2 and O_2. So, there is a non-intentional analysis for the fact that R_1 and R_2 are about O_1 and O_2 respectively.

But say R_3 and O_3 don't stand in the relation T. So, the fact that R_3 is about O_3 cannot be analyzed as R_3 and O_3 standing in the non-intentional relation T. The reductionists don't think that this is a problem at all because R_3 is composed of R_1 and R_2. So, how are we supposed to analyze *the fact that R_3 is about O_3* in light of the semantic compositional nature of R_3? Suppose we analyze the fact that R_3 is about O_3 as the following complex fact:

[**Analysis#1**] (i) R_1 stands in T with O_1; (ii) R_2 stands in T with O_2; (iii) what R_3 represents is fully determined by what R_1 and R_2 stand in T with; and (iv) (i) and (ii) determines that R_3 represents O_3.

This purports to analyze the fact that R_3 is about O_3 in terms of the aboutness of simple mental representations T plus semantic compositionality, namely, the fact that whatever R_3 represents is fixed by what R_1 and R_2 represent (i.e., what R_1 and R_2 stand in the relation T with). But [Analysis#1] is a disaster: (iii) and (iv) are pregnant with intentional notions, which undermines the project of getting intentionality out of the picture. (iv) even explicitly mentions the fact that R_3 represents O_3 — a fact that we are supposed to analyze in the first place. But taking (iii) and (iv) out and leaving (i) and (ii) alone certainly does not fully analyze the fact that R_3 refers to O_3. R_3 doesn't even feature in (i) and (ii). (Notice that the intentional notions pertaining to R_3 cannot be replaced with the non-intentional notions T because this toy example is meant to simulate one of those tricky cases where a (complex) mental representation cannot be captured directly by a naturalistic tracking relation.)

The failure of [Analysis#1] is telling. It reveals a deep obstacle in invoking semantic compositionality to reduce intentionality. Semantic compositionality is explicitly fleshed out in intentional terms: it's about how the *representational contents* of simple mental representations jointly determine the *representational contents* of complex mental representations. If semantic compositionality is invoked in our analysis, the resultant analysis wouldn't be a non-intentional reduction.[4]

Semantic compositionality should at best be considered a formal constraint on the naturalistic analysis we offer: Whatever naturalistic analysis we offer should

---

[4]It has been suggested to me that my argument here must somehow be wrong because the two-step strategy for naturalization is just an instance of a recursive definition, which must be legitimate and non-question-begging. That's a misunderstanding of recursive definition, I fear. A recursive definition isn't just whatever definition that consists of building up complex entities from simpler ones. Informally speaking, to define H(x) recursively is to assign objects into H(x)'s extension step by step, by means of building from simple objects up to complex ones (e.g., that's how the well-formed-ness of logical strings is defined). In the task of naturalizing mental representation, we aren't trying to define or analyze any single function or property. In particular, we aren't trying to analyze the abstract meaningfulness of all mental representations, we're interested in naturalizing specifically things like the fact that R_1 is about O_1, the fact that R_2 is about O_2, and the fact that R_3 is about O_3. Hence, the idea of recursive definition is irrelevant.

end up conforming with the constraint of semantic compositionality.[5] But semantic compositionality itself shouldn't be a constitutive part of a reductive analysis of intentionality. If semantic compositionality is a fact, the reductionists need to find a way to reduce that fact into certain non-intentional fact too. But if semantic compositionality can be reduced into non-intentional facts, we shouldn't need semantic compositionality to have an adequate reduction of intentionality. So, if one doesn't already have a naturalistic analysis of mental representation generally, appealing to semantic compositionality has little to offer.[6]

3 Invoking Syntactic Compositionality

How about the weaker notion of syntactic compositionality? Let's use the same example to try this out. As a first past, suppose we say the fact that R_3 is about O_3 just is the following complex fact:

> [**Analysis#2**] (i) R_1 stands in the relation T with O_1; (ii) R_2 stands in the relation T with O_2; and (iii) R_3 is syntactically composed of R_1 and R_2.

(iii) says that R_1 and R_2, as non-intentionally individuated internal states, stand in certain non-intentional relation to compose R_3, as a non-intentionally individuated internal state. [Analysis#2] refers to the fact that R_1 and R_2 compose R_3. But there is no mentioning of semantic compositionality.

The analysis is inadequate. Say I have two cards: the word "red" is written on one of them, and there is a drawing of a car on the other. The first card represents redness and the second one represents cars. Now I stick the two cards together. Do I get a composite entity that represents red cars? It depends. The cards have to be put together ***in a right way***. [7]The same goes for syntactically composed internal states. Suppose R_1 and R_2 jointly compose R_3 syntactically. That doesn't immediately give us the fact that R_3 is about O_3. The composition has to be done ***in a right way***; so let us modify the analysis:

> [**Analysis#3**] (i) R_1 stands in the relation T with O_1; (ii) R_2 stands in the relation T with O_2; and (iii) R_3 is syntactically composed of R_1 and R_2 in a right way.

---

[5]For example, Fodor (2001) uses that as a way to rule out the language-first approach to representational content. Robins (2002) is skeptical that compositionality can do even that.

[6]We shouldn't be complacent with claims like the simple mental representations are reduced to the physical and the complex ones are reduced to the simple ones. Simply asserting that something is reduced isn't enough. This paper is an attempt to press reductionists to spell out a bit more how the so-called reduction is supposed to be done. Of course, one can believe that semantic contents of complex mental representations supervene on the semantic contents of the simple mental representations. But that at best gives one metaphysical dependence, not reduction of intentionality. Note that even a substance dualist is free to accept that the mental substance is somehow metaphysically dependent on the physical. So, metaphysical dependence is way too cheap for reductionism about mental representation.

[7]Notice that I didn't say in *the* right way.

The question is of course: What counts as a right way? By exploring some potential answers to this question, I'll show that a plausible analysis of what counts as a right way inevitably contains the notion of mental representation we seek to reduce, making syntactic compositionality a dead end for the reductionists too.

To begin, let's consider the card example again. Does sticking the two cards together produce a representation for red cars? It certainly doesn't help answer the question that I simply say that the two cards are stuck together by glue. Here is a thought: that glue is involved alone doesn't help unless the glue between the two cards somehow represents an instantiation relation between whatever the cards represent individually (i.e., redness and cars).

Similarly, not only does a reductive analysis need more detail about the syntactic composition of R_3 from R_1 and R_2, the syntactic compositional feature of R_3 must somehow represent whatever relation O_3 stands in with O_1 and O_2. Only then can the syntactic compositional feature of R_3 have a plausible claim to deliver us the fact that R_3 is about O_3.

So, we have hope for a successful reduction of intentionality only if there is a way to naturalistically analyze the fact that a complex mental representation's syntactic compositional feature (e.g., the way in which R_3 is non-intentionally composed of R_1 and R_2) represents the objectual relation between the complex mental representation's object and whatever objects the component mental representations represent (e.g., the relation O_3 stands in with O_1 and O_2).[8] This is simpler said than done.

Suppose N is the *syntactic* relation that ties R_1 and R_2 together to form R_3 so that $N(R_1, R_2, R_3)$ is true; and suppose that O_3 stands in the relevant *objectual* relation M with O_1 and O_2 so that $M(O_1, O_2, O_3)$ is true. For N to be a right way to stick R_1 and R_2 together to form an R_3 that is about O_3, we need the fact $N(R_1, R_2, R_3)$ to *non-intentionally track* the fact $M(O_1, O_2, O_3)$ *in virtue of* the fact that R_1 and R_2 non-intentionally track O_1 and O_2 respectively. Here's a suggestion. We flesh out (iii) of [Analysis#3] into (iii) - (v) of [Analysis#4]:

> **[Analysis#4]** (i) $T(R_1, O_1)$; (ii) $T(R_2, O_2)$; (iii) $N(R_1, R_2, R_3)$; (iv) $M(O_1, O_2, O_3)$; and (v) for any mental states u, v, and w, and anything x, y, z, the fact that $N(u, v, w)$ stands in the relation T with the fact $M(x, y, z)$ in virtue of the fact that $T(u, x)$ and $T(v, y)$.

The clause (v) intends to capture the idea that R_1 and R_2 are syntactically put together in a way such that the syntactic compositional feature of R_3 (i.e., N) tracks (by the relation T) the relation M that links up the object of R_3 and whatever objects R_3's components (i.e., R_1 and R_2) track individually (by the

---

[8]In the context of Millikan's view, this requirement boils down to the need for a fully naturalistic rendering of the "mapping rules", which will help explain how transformation and re-combination of the components in a mental representation can produce new complex mental representations, i.e., novel thoughts.

relation T). Since T is assumed to be non-semantic, [Analysis#4] does not contain any semantic notion.

Here's [Analysis#4]'s problem. Although it works fine for complex mental representation formed by simple mental representations, it leaves no room for complex mental representations to be combined *in the same way* to form even more complex mental representations. [Analysis#4] would fail to analyze the fact that R_3 is about O_3 if either R_1 or R_2 is complex to begin with. This is due to the part "T(u, x) and T(v, y)" in clause (v).

That part is in (v) because we need to capture the idea that the syntactical feature of the complex mental representation tracks certain objectual relation that links up *whatever the component mental representations track*. Recycling a previous analogy to illustrate: Using glue to stick a card that represents redness and a card that represents cars together produces a composite object that represents red cars only if the glue we use to stick two cards together somehow represents an instantiation relation between *whatever property the first card represents* and *whatever thing the second card represents*. Otherwise, sticking two cards together with glue wouldn't create a representation of red cars no matter what the cards represent individually. What (v) in [Analysis#4] expresses is exactly what is being said about the glue.

Whereas T is the tracking relation that, we assume, adequately naturalizes simple mental representations, complex mental representations don't stand in the relation T with their objects. (That is why the reductionists appeal to compositionality in the first place.) Hence, if we use T to flesh out the idea "whatever the component mental representations track", that works only when the component mental representations are simple mental representations. Surely this limitation is problematic.

To rectify this problem so that the syntactic relation N tracks the objectual relation M in virtue of what the simpler (but not necessarily simple) mental representations track non-intentionally, we must change the (v) in [Analysis#4] into (v'):

> [**Analysis#5**] (i) $T(R_1, O_1)$; (ii) $T(R_2, O_2)$; (iii) $N(R_1, R_2, R_3)$; (iv) $M(O_1, O_2, O_3)$; and (v') for any mental states u, v, and w, and anything x, y, z, the fact that N(u, v, w) stands in the relation T with the fact M(x, y, z) in virtue of the fact that **R**(u, x) and **R**(v, y).

Where **R** stands for the *general* non-intentional tracking relation between a mental representation (regardless of whether it is complex or simple) and its object. By spelling out "whatever the component mental representations track" in terms of R instead of T, we allow those component mental representations to be either simple or complex. This is the only way to make sure that "a right way" to put mental representations together to form more complex mental representation applies not only to putting simple mental representations together, but also applies

to putting already complex mental representations together to form even more complex representations.

If we have found a general non-intentional tracking relation between a mental representation (regardless of whether it is complex or simple) and the represented, we would have our naturalistic analysis of mental representation already — mental representation just is the relation R. But that's exactly what we don't have and hope to obtain by appealing to (syntactic) compositionality in the first place. We find ourselves in a blind alley again.

The technical failure of [Analysis#4] and [Analysis#5] highlights a conceptual obstacle in using syntactic compositionality as a path to a full naturalization of mental representation. Not just any way of putting two mental representations together gives us a new complex mental representation. The composition must be done in proper ways. Spelling out what it takes for certain syntactical compositional structure to be proper inevitably *requires sensitivity to whatever the* **component** *mental representations represent*. And these component mental representations can themselves be complex or simple. As the failure of [Analysis#4] and [Analysis#5] shows, that means a naturalistic analysis of the notion of a *proper* syntactic composition requires a general and naturalistic analysis of mental representation. Hence, if we don't already have a general naturalization of mental aboutness, syntactic compositionality cannot help.

4 Conclusion

The goal of this essay *isn't* to offer a knockdown argument against naturalization. It's instead to put pressure on a key assumption of the naturalization project that has for too long been considered trivial — the assumption that once we have an adequate naturalization of simple mental representations, compositionality can do the rest. Even opponents of the naturalization program appear to share the assumption by focusing exclusively on the simple mental representations.

It seems that there are two ways the assumption can be interpreted: Either it appeals to semantic compositionality, or it appeals to syntactic compositionality. The former cannot help because semantic compositionality is itself a semantic fact that must be reduced by the light of the reductionists. The latter cannot help either because, to spell out the notion of a proper syntactic composition non-intentionally, we need a complete naturalization of mental representation already.

Perhaps there is yet another notion of compositionality that can do the magic. None of what I've said rule that out. But then the reductionists must do the work to spell that out to assure us that all we need is naturalization of the simple mental representation. There is no shortcut. Contrary to traditional wisdom, the two-step approach shouldn't be taken for granted.

References

[1] Fred Dretske. *Knowledge and the Flow of Information.* Cambridge, Massachusetts, MIT Press, 1981.

[2] Jerry Fodor. Language, thought and compositionality. *Royal Institute of Philosophy Supplements*, 48:227–242, 2001.

[3] Jerry A Fodor. *Psychosemantics: The problem of meaning in the philosophy of mind*, volume 2. MIT press, 1987.

[4] Jerry A Fodor. *A theory of content and other essays.* MIT press, 1990.

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

[6] David Papineau. *Representation and reality.* New York: Basil Blackwell, 1987.

[7] Philip Robbins. How to blunt the sword of compositionality. *Nous*, 36(2):313–334, 2002.

[8] Dan Ryder. Sinbad neurosemantics: A theory of mental representation. *Mind & Language*, 19(2):211–240, 2004.