

# Holism and Compositionality

Francis Jeffry Pelletier

## 1 Introduction

A brief acquaintance with the web shows that the terms ‘holism’<sup>1</sup> and ‘compositionality’ are used in many different fields, both academic and non-academic. And given any two academic fields, it is usually not obvious that the terms mean the same to the practitioners. Even just within philosophy and linguistics it seems that rather different conceptions are in play when we read any two different authors. We start, therefore, with a brief survey of some of the senses in which these terms are used. A little later we will make some more careful remarks about the notions.

### 1.1 Two Kinds of Compositionality

A recent interdisciplinary conference revealed that there are (at least) two different things that are thought of when the issue of compositionality is considered. One concerns a (generalized) notion of “what is a complex item (of my theory) made of?” Here are a few places where this notion arose:

PROTOTYPES: Current prototype theory says that a prototype is a structure that has a number of attribute-value pairs. A question of interest in this area is: Given that prototype-1 has a structure [A: value-a; B: value-b; . . .] and that prototype-2 has the structure [C: value-c; D: value-d . . .], is the prototype of the “combined prototypes” (that is, the conceptual combination of the two prototypes) made up only from the material in the two component prototypes, that is, from the attribute-value pairs that are in one or the other of the prototypes?

---

<sup>1</sup>‘Whole’, and presumably also ‘hole’, is derived from the Greek *ὅλος*. As a word first formed in the English language (after the 15th century, when *wh* and *h* were separated), ‘holism’ ought to be the belief in holes, perhaps along the lines argued for and against in Lewis and Lewis (1970, 1996); Casati and Varzi (1994) — or maybe the belief in holiness. But that’s not the way it is. The Oxford English Dictionary cites Smuts (1926) as the originator of this term. In this work ‘holism’ was defined as “the theory which makes the existence of ‘wholes’ a fundamental feature of the world. It regards natural objects, both animate and inanimate, as wholes and not merely as assemblages of elements or parts.” Evolutionary forces act on species, rather than individuals, with the consequence that “evolution aims at more perfect wholes”. The OED also cites Boodin (1939) is the originator of ‘wholism’: “Two conceptions. . . namely, creative synthesis (or emergence), and wholism (or gestaltism). . .”.

The discussion at the conference said ‘yes’ to this question, characterizing this as the *compositional* view of prototypes. As an example, it was said that the prototype of PET FISH *was* made up compositionally from the prototypes of PET and FISH because all the “relevant material” could be found in the these smaller prototypes.

EXPLANATION: Given hypothesis-1 that explains phenomenon-1 and hypothesis-2 that explains phenomenon-2, is the best explanation of a phenomenon that is intuitively some mixture or combination of phenomena-1-and-2 to be constructed only from the elements of hypotheses-1-and-2? If the answer is ‘yes’, then it is a compositional explanation.

NEURAL ACTIVITY: Suppose assemblage-1 of neurons is active during task-1 and assemblage-2 is active during task-2. Now consider a “supertask” which intuitively involves doing task-1 and task-2 as parts. Then: is the assemblage of neurons involved in this supertask made up out of, and only of, those neurons involved in either task-1 or task-2? Or does the new supertask bring an entirely new assemblage of neurons to bear on the new task? The claim by the authors was that, at least after the new supertask becomes a learned, repetitive action, there is a new group of neurons responsible. And thus, motor neural accounts are *not* compositional.

Underlying this type of compositionality is the slogan:

**Definition 1.** *A whole in a compositional system built up from materials in the parts.*

And accordingly, we call this “the building block version of compositionality.”

A second view of compositionality comes (mostly) from linguistic semantics.<sup>2</sup> In this conception the question is whether there is a certain relationship that holds amongst the *properties* of members of a structure. The background presumption is that the structure is compositional (in the sense just given in Definition 1 to the term); this is usually put in terms of a “part of” relation — for instance in the language case, we might be given in advance some notion of “syntactic part”, so that we know what are the syntactic parts of a whole in some structure. The question now is whether some property of arbitrary members of this structure can be defined in terms of the possession of the same type of property by the parts of that member, together with information about how these parts are syntactically combined. In the language case, the property that is usually of interest is *the meaning* of arbitrary members of the syntactically-given structure. And the question is whether there is a way to define this property — the meaning — of all complex items solely in terms of the meanings of the syntactically-given parts of the complex item and

---

<sup>2</sup>The first occurrence of the word ‘compositionality’ more or less in the sense under discussion here seems to be in Katz and Fodor (1963), although in this paper those authors appear to be most interested in some sort of “lexical composing”.

their syntactic method of combination. A ‘yes’ answer to this for every complex member signals that the semantic theory is said to assign meanings compositionally. One might note that in this kind of compositional theory, there is a second structure (in the language case, it would be “the meanings”) which is a kind of mirror of the first (syntactic) structure. So another way of asking whether a semantic theory is compositional in this sense is to ask whether there is a homomorphic mapping between the syntactic structures and the semantic structures.

In the language case, this mapping is called “the meaning function”, which I will symbolize  $\mu$ . So,  $X = \mu(A)$  means that (i) A is some element of the first structure (e.g., A is some syntactic item), and (ii) that X is the item in the meaning structure that is paired up with A (i.e., X is the meaning of A). Then this second conception of compositionality asserts that, there is a function  $\mathbf{f}$  such that whenever A is composed (in the sense of Definition 1) of B, C, D ... by means of syntactic method X, then  $\mu(A)$  is  $\mathbf{f}$  applied to  $\mu(B), \mu(C), \mu(D), \dots \mu(X)$ . That is: the system is compositional if and only if, there is a function  $\mathbf{f}$  such that for every (syntactically) complex item A in the syntactic system, its meaning  $\mu(A)$  is a function of, and only of, the meanings of A’s syntactic parts, together with the way they are combined. If  $NP_1$  and  $VP_1$  make up the entirety of  $S_1$  and they are combined by rule-X to do so, then  $\mu(S_1) = \mathbf{f}(\mu(NP_1), \mu(VP_1), \mu(X))$ . Underlying this type of compositionality is the slogan:

**Definition 2.** *The  $\mu$  of a whole is a function of the  $\mu$ ’s of its parts and the ways those parts are combined.*

And accordingly, we call this “the functional version of compositionality”.

A difference between the two notions of compositionality concerns whether some “whole” can contain things not in the parts. According to the building-block view, no; but according to the functional version, yes. For, the first notion allows the whole to contain only what is in the parts, possibly re-arranged in some manner. But the second allows the thing associated with a whole (in the linguistic case: the meaning of a complex whole) to be a *function* of the things associated with the parts (in the linguistic case: a function of the meanings of the syntactic parts and syntactic mode of combination). There is nothing to stop such a function from introducing new material into the thing associated with the whole — that is, the function can make the meaning of a whole contain many new and radically different things than are contained in the meanings of the parts. Indeed, it could introduce *all* new material, and contain *none* of the properties of the subparts. According to Definition 2, all that is required is that this be a function — which merely means that it must introduce this same material every time it is faced with the same parts and manner of combination. In the case of describing the neurons active in the complex task, the function  $\mathbf{f}$  need not pick out any of the neurons that are active in the subtasks ... but it would still be compositional.

This means that one cannot deny compositionality of this second sort merely with the claim that some whole is “more than the sum of its parts”; for, even if that is true, we

haven't thereby denied that there is some function that can define the whole from the parts. (This is the result of confusing the two types of compositionality. In the first type, the "building block" theory, the only things in the whole are re-arrangements of the building blocks. But this does not apply to compositionality of the second sort, wherein there might be a function that transmogrifies properties of building blocks into some entirely different things or stuff.)

Now, one might argue that the distinction between the two conceptions of compositionality merely reflects a difference between "ontological compositionality" and "semantic compositionality". But a closer inspection shows that those proposing the differing conceptions do not view it in that way. The prototype example, for instance, shows that those with interests in "the meaning of mental items" can use the 'building block' notion. And of course, the prototype conception of meaning is rife within some linguistic theories, such as those advocated by Ron Langacker, Len Talmy, and perhaps Ray Jackendoff.<sup>3</sup> Various other authors have proposed still different theories of meaning that are compositional in the first sense. Using the notion of a 'semantic differential', Osgood et al. (1957) proposes a finite number of "dimensions of meaning", and every lexical item is said to have some value along each dimension. For example, along a *pleasant—unpleasant* dimension, each lexical item receives some numerical value; along an *active—passive* dimension they each receives some numerical value; and so on. All meanings of more complex items are seen as being constructed from the values, and they too will have their own vector of meaning defined by their [dimension:value] set of pairs. In general, *any* theory that postulates a finite vocabulary of primitives and a finite number of ways to evaluate a primitive will be compositional in this first sense so long as the ways of combining the part-values is unique — that is, there is no way to get two different complexes when given the same parts combined the same way. E.g., the notion of universal (cross-cultural) semantic primitives is like this (Wierzbicka, 1996), as is the notion of basic conceptual dependencies (Schank, 1972), and most semantic decomposition models, e.g., Jackendoff (1990) among others. The idea is that all one can use in constructing a complex will already be present in the parts. The construction can of course differ if the parts are "put together" differently, just as a combination of two building blocks differs depending on whether they are laid side-by-side or one-atop-the other. (For the example of combining prototypes, a building-block compositional theory distinguishes 'flea circus' from 'circus flea', 'gun grease' from 'grease gun', 'school teacher' from 'teacher school', and so forth, in this manner.)

So, I do not think it useful to distinguish the two notions of compositionality as merely being due to applying the same concept in different areas of inquiry. Instead, there just are two very different notions, of course with some overlap in conception, that yield very different accounts of and rationales for views about how complexes are to be understood.

---

<sup>3</sup>See, for example, (Langacker, 1987; Talmy, 2000; Jackendoff, 2002), even though these theorists don't all think their theories are compositional.

## 1.2 Two Kinds of Holism

The same conference that revealed a difference in understanding of the term ‘compositionality’ also displayed a basic difference between two ways of understanding ‘holism’. The fundamental idea behind holism is to give some sort of priority to a “whole” in preference to the “parts” of that whole, and of course this plays out differently in different areas where the term is used. But there is a basic difference between ‘holism’ as describing the view that there are objects (“wholes”) that need to be referenced in some realm of inquiry, and ‘holism’ as describing the view that properties of the individuals under consideration require reference to properties of other individuals before the original properties can be fully defined or explained. The first sort of holism concerns what entities have basic existence:

**Definition 3.** *Some properties can only be attributed to entities that are not individuals.*

and accordingly, we call this “ontological holism.”

Ontological holism is raised in very many fields of enquiry, especially in the social sciences and humanities. Businesses and corporations, for instance, might have duties and obligations (etc.) that are not duties or obligations of any individual within the business. They might similarly have legal rights and legal constraints that do not devolve to any individual. Nations might have properties that are not properties of any members or collections of members of the nation. They can declare war, or decree that their borders be closed; but no individual in the nation can do so, not even the leader. Thus there must be these entities, call them “wholes”, that are distinct from, and not “reducible to”, the set of individuals that make them up.

Something similar to this is sometimes postulated within some sciences. Such a scientific holism is the view that some scientific laws must make reference to populations/species/etc. The type of example that is often given<sup>4</sup> concerns the interaction of two populations:

Suppose we have an ecological system composed of seals and cod. There are periodic fluctuations in their population levels. Seals eat cod to the point that there are too few cod to sustain the seal population, so they begin to die out. But this allows the cod to multiply. And in turn this provides food enough for the seal population so they can multiply ... and so on, cyclically.

The claim is that this is a type of scientific law that cannot be rephrased in terms of individual seals and cod, but makes essential reference to the notion of a “population” or (perhaps) to species. Quantum holism is another area that introduces wholes of this type. Here, the view is that an “entangled pair” of quantum particles form a unity that is distinct from the individual particles that constitute the “whole”. Further issues concerning compositionality in science come from consideration of theoretical terms, which seem to be

---

<sup>4</sup>For example, by Garfinkel (1981, Chap. 2).

holistic in that they are defined only by the axioms of the theory that introduce them. For discussion, see Schurz (2005, esp. §5).

Contrasted with ontological holism is talk about properties that are somehow interconnected:

**Definition 4.** *Some properties of an object are defined in terms of the same type of property of some other object(s), and these properties of the other object(s) are in turn defined by means of the first property.*

and accordingly, we call this “property holism.”

The “partiality” inherent in this definition of property holism is difficult to overcome, for proponents of property holism do not believe that there are any “independent” properties from which one can start and thereby develop an account of all the holistically-related properties. And similarly, the appearance of circularity is mandated by the fact that such holists do not think that any of these properties can be described independently of any others. For, they would say, this is just what it means to be holistic. (We return to this issue of “the direction of explanation” below, in the next section.)

Definition 4 could describe just a very small holistic system containing but two objects and some property they each have. But of course the main cases of interest are where there are very many such objects and they each have a property that is defined in terms of that same category of property as manifested by all the other objects that make up this group. As examples, perhaps ‘beautiful’ is (partially) defined by reference to the meaning of ‘ugly’, which in turn is (partially) defined by the meaning of ‘beautiful’. Or, it might be claimed that the meaning of ‘aggression’ is (partially) defined in terms of the meaning of ‘defense’ and (partially) in terms of the meaning of ‘instigate’ and partially in terms of other words of this nature. They in turn are defined (partially) by the word ‘aggression’.

A natural conclusion to draw from Property Holism is that a “whole” is formed from the items which are related by having some property of each of them be defined in terms of the others. That is, a whole of the sort envisaged by Definition 3. Although this might be a natural conclusion, I think it is not inevitable. For example, the letter of Definition 4 makes the properties remain as properties of the individual items, and does not demand that they be attributed to any yet further whole, as would be required by Definition 3. Keeping these two notions of holism separate may shield the believers in one or the other types of holism from criticisms leveled at the other type, in much the same way that keeping the two notions of compositionality separate can protect the believers of one type from criticisms leveled at the other type.

It can also be noted that the two distinctions — one within compositionality and the other within holism — are really quite similar. Definition 1 in effect denies the (“primary”) existence of wholes; Definition 3 affirms it. Definition 2 claims that properties of wholes can be functionally computed from the similar properties of their parts; Definition 4 denies this. The Definition 1 vs. Definition 3 dispute is about the existence of (kinds of) entities; the Definition 2 vs. Definition 4 dispute concerns how properties of objects are to be

understood or defined. The Definition 1 vs. Definition 3 dispute is the age-old conflict between *atomism* vs. *wholism* (with a ‘w’).<sup>5</sup> The Definition 2 vs. Definition 4 dispute is a newer conflict about how properties of objects are best explained, and this dispute is usually called *compositionality* vs. *holism*. And while there are connections between the atomism and compositionality, and other connections between wholism and holism, it does none of the four positions good to have it conflated with some other position to which they are not committed.<sup>6</sup> So, I will use the four names ‘atomism’, ‘wholism’, ‘compositionality’, and ‘holism’ for the remainder of the paper.

## 2 Atomism and Wholism

The dispute between atomism and wholism is fundamentally ontological: Are there non-individual items in the world? Of course, such a dispute turns crucially on what is meant by ‘(non-)individual’, and so it is rather surprising to find that many of the disputants agree on what is meant, or at least, do not seem to challenge each other about this. The kinds of argumentation in this dispute instead ranges from the explanatory abilities of (natural and social) scientific laws to psychological phenomena to issues in “the direction of explanation.”

As part of their belief in the existence of wholes that cannot be defined in terms of their parts (“the whole is more than the sum of its parts”) — primitive wholes, let’s call them — wholists also believe

- (1) A primitive whole determines the nature of its parts
- (2) The parts cannot be understood when considered in isolation from the primitive whole of which it is a part

And they usually also add

- (3) The parts of a primitive whole are dynamically interrelated and interdependent
- (4) Analysis as a methodology fails in the case of primitive wholes

And conversely, atomists deny these four claims.

Since the central topic of this paper is the compositionality-holism dispute, we cannot go deeply into a discussion of the atomism-wholism dispute. However, it seems apposite to say a few words nonetheless, since there are people who confuse the considerations relevant only to this issue with those relevant to the other debate. For instance, compositionists

---

<sup>5</sup>Yes, there is also a quasi-religious movement called ‘wholism’, not to be confused with what followers of Definition 3 believe.

<sup>6</sup>For instance, there is unending argumentation between atomists and wholists over the correctness and applicability of “reduction”. But this sort of consideration is just beside the point in the dispute between compositionists and holists.

will admit of wholes that are not defined entirely in terms of their parts. In the case of language, for example, they hold that sentences (and other syntactic units) are more than a set of words — sentences are wholes that take a set of words and form a unity by blending them together in accordance with some syntactic rules. And this is a different sort of entity — a whole — that is not to be identified with the set of words that comprise it. The various wholes that are formed in compositional systems of language *cannot* be identified with the set of the words comprising the whole; that is why the formulations of compositionality in language always look like “the meanings of its parts *and* the way these parts are combined”. Nonetheless, when arguing against compositionality, wholists/holists will claim that wholes cannot be identified with their parts, and they will give examples from many different fields where this alleged to be clearly true. But as an argument against compositionality in language, this claim just misses the mark, since the compositionists *are* committed to their own sorts of wholes. Their differences with holism lie elsewhere.

Sometimes wholists bring forward considerations of gestalt theory and emergent properties to bolster the claim that there are primitive wholes. (Recall the earlier quotation from Boodin 1939, who identified wholism and gestaltism). The square that is formed when very heavily-blackened cut-out corners are drawn is one such gestalt figure that is common in the literature. “Surely,” it is claimed, “the square is *not at all* to be analyzed as those four heavily-blackened cut-out corners.” But equally surely, the atomist never claimed any such thing. Rather, the square requires that the corners be placed *exactly so* in relation to one another to make the square appear. The square could be claimed by atomists to be a function of the heavy corners plus their method of arrangement or combination.

Similar remarks can be made about emergent properties, that is, properties that the wholes have that are not possessed by any of the parts. For example, hydrogen and oxygen combining to give off energy and to produce water. The claim of wholists is that such properties are not “sums” of the properties of the initial hydrogen and oxygen. And since they are not “sums” of the properties of the initial parts, it has seemed to some that this calls for a new whole that possesses this property. But we can see that such considerations do not really play a role here. An atomist would attribute the property to the *juxtaposition* of the hydrogen and oxygen (plus whatever else is needed to initiate the reaction).<sup>7</sup>

According to points (1) and (2), the whole *determines* the nature of its parts, a doctrine which is interpreted by wholists as entailing that the whole determines the properties that the parts have. Phillips (1976, p.8) puts the view’s rationale like this:

In effect, [Bradley] maintained that when entity A enters into a relationship with entity B or C, it gains some property or characteristic, *p*, as a result of this relationship. Without the relationship, and hence without property *p*,

---

<sup>7</sup>In fact, historically this seems to be the initial purpose in introducing emergent properties — to avoid having to postulate some new force that operates “on a different level from” the ordinary objects. This was pretty successful in eliminating “the vital force” from accounts of life in the late 1800s and early 1900s. See McLaughlin (1992) for an historical account along these lines, but also see O’Connor and Wong (2006) for an opposing view of the history of emergentism.

Bradley argues, A would be different, or not-A. Any relation at all between A and another entity necessarily determines some property of A, without which A would not be what it is. This is the heart of the theory of internal relations: entities are *necessarily* altered by the relations into which they enter.

When some entities form a whole, this theory claims that their defining characteristics will be determined at least in part by the relational properties they thus possess. If the whole were different by containing some further item, then the old members would themselves be different in their essential qualities. So, parts cannot be understood in isolation from the whole in which they are parts.

So far as points (1) and (2) go, Phillips (1976) says that these claims make sense only against the backdrop of a theory of “internal relations”. And, since Phillips also thinks that this theory is intellectually bankrupt, he concludes that these wholistic doctrines cannot stand. The doctrine of internal relations claims that (apparent) relational properties which two or more items bear to one another are actually intrinsic properties of those items. If A is taller than B, this “taller than B” is one of A’s inherent properties (and “shorter than A” is one of B’s inherent properties). The history of philosophy has not been kind to this doctrine. If it is an inherent property of A that he is 175cm tall, and an inherent property of B that she is 160cm tall, this still does not adequately characterize the tallness-in-A-and-B, for it does not tell us that the relation of taller-than is asymmetric, irreflexive, and transitive. And these properties of *the relation*, taller, cannot be captured by monadic properties of individuals. G.E. Moore and Bertrand Russell were instrumental in disposing of internal relations more than a century ago. Modern wholists seem to forget that this aspect of their position requires them to adopt a notion of internal relations.

An important feature of much reasoning by most wholists is a concentration on “failures of analysis”. It is claimed that no one could predict the behavior of a certain mass of gas, for example, looking merely at the intrinsic properties of the atoms that describe it; and thus there is a failure of analysis as a methodological principle. As Phillips (1976, p. 12) says, however, this overlooks the fact that any proponent of methodological analysis would say that the laws of the system had first to be known, and that the initial conditions of the system had to be described. In fact, a wholist will be in the same condition, will s/he not? If the “whole” is the mass of gas and its behavior, it seems implausible that anyone could predict the items that make up the whole without the laws of the system — regardless of whether they are atomists or wholists.

Some wholists are also attracted to the argument that, methodologically, one wants to eschew reductionism; and there is usually presented at this stage a number of shortcomings of reductionism. But saying that “from knowledge of the parts only one cannot predict the properties of the whole” is quite different from saying that *after* knowledge of the whole has somehow been attained, *then* this cannot be reduced to, or explained in terms of, the atomic parts. And surely this latter notion of analysis is that to which the atomists are committed.

In any case, though, it seems to me that all these arguments start from the wrong point. Isn't the atomist committed only to saying that *in reality* the wholes are sums of their parts? Surely whether or not we *know how to construct* the wholes from their parts is irrelevant to this, and doesn't at all impugn atomism. Aren't these epistemic considerations completely beside the point in the atomism-wholism dispute about ontology?

I have concentrated in this section on the ontological issues of atomism and wholism because they often seep into discussions of compositionality and holism. I hope this discussion now makes clear the sorts of considerations that are *not* relevant to our discussion in the next section, where we *will* talk about compositionality and holism. And although I have been concentrating, in this section, on the ways that an atomist can respond to the charges that wholists make in favor of their theory and against the opposed atomist theory, this shows only that the wholist has not adequately cinched his/her case. It is not thereby shown that the atomistic position is proved, and in fact I do not believe it has been. It has long seemed to me that atomism and wholism are "structurally identical" theories, differing only in starting points. Intuitively, an atomistic theory postulates a set of basic entities and defines wholes by combining these entities in certain prescribed ways, generating bigger and bigger wholes, all of which are non-primitive. A wholist postulates<sup>8</sup> an all-encompassing primitive whole and defines smaller and smaller wholes based on prescribed ways of decomposing, some of which will be primitive and others not. If it is true that these theories are structurally identical, then there is no difference between them other than their names, and what they choose to name items in their self-same theory. But a discussion of these issues must await a different venue.

### 3 Compositionality and Holism

#### 3.1 Is There a Conflict?

The definitions given earlier for compositionality and holism were:

A system is **compositional with respect to property  $\mu$**  if and only if there is a function  **$f$**  such that the  $\mu$  of any complex object in the system is that function of the  $\mu$ 's of its parts.

A system is **holistic with respect to property  $\mu$**  if and only if the  $\mu$ 's of complex objects in the system are defined partially in terms of the  $\mu$ 's of some other objects of the system, and the  $\mu$ 's of these other objects are in turn defined partially by means of the the  $\mu$ 's of the first object.

This article is for the most part concerned with the use of these notions in semantics, and so  $\mu$  is interpreted as "the meaning of", and it is in that sense that I will be concerned with compositional and holistic systems when I speak of them without further qualification.

---

<sup>8</sup>Or at least, the type of wholist I am imagining.

The systems themselves are collections of syntactically-defined objects, and the notion of “part” is therefore understood as being a syntactic part. The key notions used in defining compositional and holistic systems could use some further explication and perhaps some tidying-up. But on an intuitive level, the two types of systems are supposed to be in conflict. Thus, my explications will be aimed at making compositional and holistic semantic systems be contrary to each other, after which we will see which side of the conflict seems a more viable approach to semantics.

It will not have escaped the reader’s notice that there has been rather a lot of “slack” in the accounts offered for compositionality and holism. And therefore, many writers have been moved to show that the two are compatible. And certainly they are, if one picks the right definitions. If one focuses on the

**Slogan:** The meaning of a word is its contributions to the sentences in which it occurs.

More generally, the meaning of any expression is the contribution it makes to all the larger wholes of which it is a part.

it will then seem that the two are equivalent, for this slogan can be seen as a description of *both* compositionality and holism (especially to those holists who focus on “contextualism”). I do not here go into a discussion of how the two theories can both have this Slogan as a consequence, but readers of the literature can find numerous examples. So, this Slogan does not set compositionality apart from holism, and for this reason we do not consider it any further because it cannot show the “conflict” that is supposed to hold between the two viewpoints. At least, it is my view that the two are opposed notions in most theorists’ minds and I therefore wish to discover an interpretation of the notions that makes this happen.

Other ways of using the “play” in the terms that are used to define ‘holism’ and ‘compositionality’ to show that they are after all compatible can be found in discussions of Frege. When addressing their beliefs that Frege both held to compositionality and to holism (the latter in virtue of his “Don’t ask for the meaning of a word except in the context of a sentence” (1884, §60)), scholars such as Dummett, Currie, Haaparanta, Hintikka,<sup>9</sup> have gone to extraordinary lengths to show that Frege was not being simply inconsistent. In this same vein, although not in the context of Fregean exegesis, Pagin (1997, p.13) is concerned to find a “*reasonable* interpretation of [Holism]. . . which is compatible with Compositionality.”<sup>10</sup>

My goal here is rather the opposite: to show why holism and compositionality are opposed theories. I take interpretations of the terms that make them be compatible as evidence against these interpretations. I want to know why the issue of holism vs. compositionality actually divides theorists of language; I’m not interested in explanations to the effect that “you’re both right, just about different things.” Of course there might be very many interpretations that make them out to be incompatible, just as there are many

---

<sup>9</sup>See Dummett (1981a,b); Currie (1982); Hintikka (1984); Haaparanta (1985); Baker and Hacker (1980). These (and many others) are discussed in Pelletier (2000a).

<sup>10</sup>Although on most other issues I am in broad agreement with Pagin (1997, 2006).

interpretations that make them be compatible, but I will focus on what is offered in the literature. Once we can come up with an interpretation of this nature, we will then be in a position to evaluate compositionality as against holism.

### 3.2 Further Characterization of Compositionality

Many scholars think that the notion of compositionality is much clearer than that of holism, and perhaps they are right. But there are nonetheless many things about Definition 2 that could stand to be more carefully delineated. The issues fall generally into three categories: What is it to be a function of the sort under consideration? What are meanings and their parts? What restrictions are there on the underlying syntax? There is the further, very general issue concerning the idea that compositionality, as opposed to holism, is supposed to be a theory of *local* meaning. But the characterization of compositionality in Definition 2 says only that there is *some mathematical function* that relates the parts to the more complex. In itself, this does not build in any notion of locality vs. globality; that was, in fact, one of the morals to be drawn from the earlier discussion of how some characterizations of compositionality can be made compatible with holism. Once a system is characterized in whatever way, then there can very often be a mathematical function that relates the complex items with the simple ones; but that does not explain the intuitive notion of *depends upon*, which is supposed, in turn, to tell us about *the order of explanation* inherent in compositionality, and which is supposed to be different in holism.

So, I wish to understand this “functional relationship” in an ontological way. The function is supposed to correspond to the way that the meaning of complexes *really depend* on the meanings of their parts. Even adopting this understanding of function, there are some things that ought to be cleared up in our talk about functions. When one says that X is a function of Y, they might be talking either mathematically or informally. In the latter case they just mean that in one way or another Y is important for X. I am presuming that our compositional theorists are not talking informally, but rather that they mean their function to be one in the mathematical sense, augmented by a notion of “depends on”. I believe that the intent of compositionality is to say that the meaning of the complex item is a function of, *and only of*, the meanings of its parts and manner of composition.<sup>11</sup> But even if we tighten up in this way, then there can be unwanted cases, some of which are surveyed in Pelletier (1994a). I am presuming here that these niceties are taken care of.

---

<sup>11</sup>Hintikka (1980); Partee (1984) both think, contrary to what I have just said, that the common intent of compositionality is to allow factors other than the meaning of the parts as possible influences on the meaning of the complex. Hintikka calls having the meaning of all wholes be completely determined by the meanings of the parts the *determinacy thesis*. Partee thinks that there are a number of cases that cast doubt on the determinacy thesis. It seems to me that compositionality has *always* had the determinacy thesis in mind when advocating compositionality. After all, *opponents* argue *against* compositionality by trying to show that there are *some* features other than the parts which are relevant; they do not think they have to show that the parts *never* play a role.

Some disputes about compositionality turn on what is meant by ‘meaning.’ For instance, if the syntactic form of the part is considered to be an aspect of the meaning of the part, then it can be shown that “any semantics can be made to be compositional”.<sup>12</sup> Thus, theories that take the meaning of a word to be the pair consisting of the actual word and its word-meaning as normally understood, and use this information to construct the meaning of a next-most-complex unit, which meaning is also a pair consisting of this syntactic description of the item plus a second element computed from the word-meanings, are susceptible to the charge of triviality. Westerståhl (1998) considers other proofs that purport to have much the same upshot — that compositionality is “trivial” or “a methodological matter” (meaning that it is merely a theoretician’s preference to work with a theory of that nature than with a theory of some other nature, but which has no empirical consequences). Westerståhl brings out the sort of presuppositions these proofs make, and argues that they are all implausible assumptions.

One direction not considered by Westerståhl, but which seems related to theories that build the syntactic form into the meaning (a kind of “structured meaning” approach, perhaps), are theories that want meaning functions to operate not only on the meanings of the parts, but also on the parts of these parts, and so on. Thus, the meaning function takes as arguments not only the *immediate* syntactic constituents, but on their sub-constituents. This distinction is mentioned in Partee (1984), and in Larson and Segal (1995) it is called the difference between “strictly local” and “non-local” compositionality. To see the difference, suppose the meaning of some node in a syntax tree is a set of possible worlds.<sup>13</sup> A strictly local compositional semantic rule would use this set of possible worlds (plus information about the way its syntactic value is combined with another meaning) to compute the meaning of the parent node. But a non-local rule could have access to information such as *which part of that set comes from which subpart of the syntactic item*, and this in turn seems very close to the above-rejected picture of meaning where the syntactic form is considered a part of the meaning.

In a similar vein, if we are given a language — a syntax plus semantics — where the semantics is non-compositional, and if one is allowed to change freely the syntactic structure assigned to the elements of the language, then the same semantic values can be assigned to the sentences (strings of words, now having a different structure) of the language as before but in a compositional way.<sup>14</sup> For example, if one is allowed to introduce two syntactic forms where there used to be just one in the original language, then (some) violations of compositionality can be given a compositional treatment. But surely this sort of change is quite different from the charge that compositionality itself is vacuous. In the case of

---

<sup>12</sup>See Westerståhl (2004, 551–552). I believe Putnam (1954) was the first to suggest this as a way to save compositionality in the face of Mates-type examples (Mates, 1950).

<sup>13</sup>Just suppose. You don’t have to believe it.

<sup>14</sup>See the discussion in Westerståhl (1998) for proofs of this and related matters. See also Westerståhl (2004) for discussion of the conditions under which a compositional partial semantics can be extended to an entire language.

creating new syntactic forms, we need to have some *syntactic* evidence for this; and if we do have it that is our evidence for a compositional treatment, and if we don't have it that is our evidence against compositionality for this language's semantics. (Of course, if the desire for compositionality is the *only* motivation for such syntactic changes in every case, then we might indeed accuse the theoretician of treating compositionality as a methodological principle. But this does not mean that the general notion of compositionality is itself vacuous, only its employment in this particular case.)

We have just seen that some versions of compositionality are vacuous because they define meaning in such a way as to encode all the lexical and syntactic information, or they allow the meaning function to have access to the way that meanings of parts was computed, or they allow otherwise unmotivated changes to the underlying syntactic part-whole relation. And as a consequence, it becomes always possible to re-code a non-compositional semantics as a compositional one. But ignoring these sorts of features and tricks, one might ask, is the whole notion still vacuous? Is it maybe just that semantic theoreticians are drawn to the “neat and tidy” picture that compositionality offers — the issue of compositionality vs. non-compositionality in semantics is perhaps just an aesthetic taste?

The answer to this is no: there do seem to be non-compositional features of natural languages. I am not going to discuss these here: they are the “argument from synonymy” and the “argument from ambiguity” discussed in Pelletier (1994a,b, 2000b). Perhaps further investigation will yield evidence that these features are best described in some other ways that do not generate non-compositionality. But this further evidence needs to be independent from the desire to retain compositionality, or else the principle truly is “merely methodological” in linguistics.

### 3.3 Further Characterization of Holism

Much of the opposition between holism and compositionality comes from considerations that are lumped together under the banner of *direction of explanation*. In a holistic system, it is the “whole” that is considered primary and it is this whole that is the “cause” or “explanation” of the meaning of the words. In a compositional system, on the other hand, the words are primary, and it is they (plus their modes of combination with other words) that “cause” or “explain” why larger units like sentences have the meaning they do.

The sort of holism that I am interested in — the sort that is *really* opposed to compositionality — is expressed in Saussure (1916). It is a complex explanation, but the various aspects of this explanation will illustrate many of the features in holism that composition-  
alists claim to be faulty. Let us start with his general characterization (Saussure, 1916, 114). He distinguishes “value”, or “content”, from “meaning”; it is the notion of value that I wish to focus on.

[A word's] value is therefore not determined merely by that concept or meaning for which it is a token. It must also be assessed against comparable values, by

contrast with other words. The content of a word is determined in the final analysis not by what it contains but by what exists outside it. . . .

The French word *mouton* may have the same meaning as the English word *sheep*; but it does not have the same value. . . . The difference in value between *sheep* and *mouton* hinges on the fact that in English there is another word *mutton* for the meat, whereas *mouton* in French covers both.

In a given language, all the words which express neighbouring ideas help define one another's meaning. Each of a set of synonyms like 'to dread', 'to fear', 'to be afraid' has its particular value only because they stand in contrast with one another. If 'to dread' did not exist, its content would be shared out among its competitors. . . . So the value of any given word is determined by what other words there are in that particular area of the vocabulary. . . . No word has a value that can be identified independently of what else there is in its vicinity.

(Since I'm using Saussure merely to give a feel for the type of view that I want to call holism, I propose to delete any reference to Saussure's notion of meaning and just talk about his notion of value — although I will feel free to call it 'meaning'.)

Like most semantic holistic theories, however, Saussure's theory focuses on the lexical items and not on sentences. On the surface — and maybe deep down inside — this seems to be a shortcoming, since pretty much all theories of language think of sentences, or even longer stretches of discourse, as defining 'language'. (Or alternatively, they think of language as the ability to generate these longer items.) Can it really be the case that a language is holistic if its lexical items are given meaning in the way indicated by Saussure? Perhaps; let's see.

The presumption apparently made by these holists who focus on the lexical interpenetration of each meaning by all other members of the lexicon is that this same interpenetration will infect sentences as well. It might be noted that **such** the semantic rules that such a holist could propose might be of a form that compositionists would like — they could combine two or more lexically interpenetrated meanings in some specified way — and the result would be nonetheless considered objectionably holistic by the compositionist's lights. For, these complexes will also have interpenetrated meanings: for example, no sentence's meaning can be understood without understanding all the various interpenetrations that this sentence has with other items. And this now means not only the rest of the lexical items, but also all the sentences in which these lexical items can occur. And that pretty much means the entire language, doesn't it? So, to understand any sentence at all, one needs to understand every sentence in the language. So, the apparent ability to combine a holistic lexicon with a compositionally-acceptable set of semantic rules does not remove the offensive feature from the holist's theory. It still exhibits what anti-holists find most objectionable: the direction of explanation and understanding.

Unlike most other semantic holist theories, Saussure's is not individualistic. Instead, he seems to be thinking of "the French lexicon in the abstract", and how its basic units

are assigned values. Most other theories start with an individual's *mental* lexicon, and argue that the way these items are acquired dictates that their meanings are all defined in terms of others in their lexicon. As we will see later, a consequence of making this holism be a matter of individuals' private lexicons is that such theories become vulnerable to the charge that interpersonal communication is impossible because each person learns their meanings differently. But a theory like Saussure's trades this difficulty in for a difficulty in understanding how one can "learn the lexicon of French" at all — another topic to which we will return.

Saussure also thinks that the claims one can make about "the French language" are identical to those we can make about any (native?) speaker of the French language. In more up-to-date terminology, he thinks that the semantic facts concerning the "external", independent-of-people, French language are identical to the "internal" semantic facts of a (native?) speaker of French. Saussure says (Saussure, 1916, 110–111):

Psychologically, setting aside its expression in words, our thought is simply a vague, shapeless mass. . . . In itself, thought is like a swirling cloud, where no shape is intrinsically determinate. No ideas are established in advance, and nothing is distinct, before the introduction of linguistic structure.

But do sounds, which lie outside this nebulous world of thought, in themselves constitute entities established in advance? No more than ideas do. The substance of sound is no more fixed or rigid than that of thought. It does not offer a ready-made mould, with shapes that thought must inevitably conform to. It is a malleable material which can be fashioned into separate parts in order to supply the signals which thought has need of. So we can envisage the linguistic phenomenon in its entirety — the language, that is — as a series of adjoining subdivisions simultaneously imprinted both on the plane of vague, amorphous thought, and on the equally featureless plane of sound. [...]

Just as it is impossible to take a pair of scissors and cut one side of paper without at the same time cutting the other, so it is impossible in a language to isolate sound from thought, or thought from sound. To separate the two for theoretical purposes takes us into either pure psychology or pure phonetics, not linguistics.

Linguistics, then, operates along this margin, where sound and thought meet.  
*The contact between them gives rise to a form, not a substance.*

Focussing on what Saussure calls the "value" of words, let's look a little more closely at what this theory would say about more modern issues. First note that the view is *not* epistemic: it is not merely that we can't *know* the value of a word without knowing the values of words that are its competitors, but that the word *doesn't have* a value without these oppositions. Saussure's holism is bound up with the view that lexical items are defined "by contrast" with others in the same neighborhood. Although Saussure talks

about such “closely related” lexical items, surely the effect can’t be so easily localized. The “semantic field” of the terms Saussure used in his example are also related to items in the “motivational semantic field”, and to those in the “moral semantic field”, and so on indefinitely. And these relations or “interpenetrations” can’t really just be “contrast”, unless this term is taken very generally indeed. Most theorists, both holists and anti-holists, would like to include *any* epistemic liaison there may be between two items. And I suspect that Saussure would want his liaisons to be broader than the “contrast” he mentions — at least, that’s the type of holism I am interested in opposing to compositionality.

As I remarked, Saussure’s view is that the “external language” is itself holistic. And the entire language, with all its sentences and other structures, is holistic because the lexical items are defined by means of liaisons with one another.<sup>15</sup> Although Saussure, like other holists, discusses only lexical items and not sentences, he nonetheless differs from the more usual take by holists, for whom the whole is restricted to an individual’s lexicon. Most holist theorists (unlike Saussure’s) do not hold that all people “tap into” the same independent-of-an-individual lexicon.<sup>16</sup> To my mind, it is no accident that most holistic theories have the feature of being lexically oriented, for there simply is no sense in the notion of a *sentence* being “admitted” into the language without any further restrictions. (One possible such further restriction is that the sentence be true, or considered true; but this move in the direction of confirmation holism does not capture *meaning* holism, in the sense that most holists want.<sup>17</sup>) The two features are related: once you see that there is no such independently specifiable whole as “all the possible meanings in the language”, then you will choose to work with the (finite) lexicon and find your holism there. But if you start there, you most naturally are attracted to how an individual acquires the lexicon, thereby making your holism be individualistic. Saussure is an exception in this last regard, and this is why I chose his view as my exemplar of semantic holism that is really at odds with compositional treatments of language.

Once one therefore fixes on lexical items, it seems that the most natural direction to take is to think of each person’s subjective lexicon as separate wholes, and deny that they are *subparts* of some larger, social whole, or somehow *manifestations* of such a larger whole. In Section 3.4.1 we will look at a case where we *do* talk of a whole as being comprised of sentence meanings. But it will not form the sort of holism we seek here.

---

<sup>15</sup>It might further be noted that the meaning of a word is *not* characterized as its contributions to all the sentences in which it occurs.

<sup>16</sup>There are some who think people can tap into “how society uses language” or “a form of life”. These do not seem any more clear than tapping into “the French lexicon” in the abstract.

<sup>17</sup>Some Davidsonians think otherwise. They think that a theory can leverage the notion of “accepted true” to escape the internalism or individualism that holism seems committed to. But I won’t follow up this strand in my discussion. I’ll discuss my own take on the issue of true sentences in Section 3.4.1, but it will not involve Davidsonian considerations. Some further Davidsonian issues will come out in Section 3.4.2.

### 3.4 Compositionality vs. Holism

Given how popular the doctrine is, it is rather surprising that there are not very many arguments that support semantic compositionality directly. Most arguments from compositionists on the general topic of holism and compositionality *assume* compositionality to be true and then go on to show that such-and-so alternative position can't be right.<sup>18</sup> In the linguistics literature there is, to be sure, some argumentation concerning the proper form of a linguistic theory; and in that milieu, semantic compositionality is praised for its clarity and transparency. But these “methodological” (aesthetic?) considerations don't really count as direct arguments in favor of the position, in my mind.

This article on holism is not the place to review the considerations that have been put forward both for and against (semantic) compositionality<sup>19</sup>, but we should look at what Jerry Fodor calls “the sword of compositionality”<sup>20</sup>, since it is usually seen as decisive against (semantic) holism.

These main considerations in favor of the semantic compositionality of language in the sense of Definition 2 come from a series of similar “arguments” that we might collectively call “The Infinity of Language”.

**[Argument from understanding]** We can *understand* an infinite number of novel sentences, so long as they employ words we already understand. We understand sentences and other combinations that we have never encountered. So, language must be ‘compositional’: it must start with a finite stock of words/morphemes and put these together in a finite number of different ways, but using an unlimited recursive method.

**[Argument from productivity/creativity]** We can *create* new sentences that we have never heard or used before, and we know that they are appropriate to the situation in which we use them. This can only happen if language is ‘compositionally’ organized, so that we learn some finite base of words and rules, but know how to combine them recursively so as to produce totally new descriptions.

**[Argument from learnability]** We are finite creatures who are exposed to a finite amount of information concerning our language. Nonetheless we can *learn* a system that is capable of infinite expression. The only way this can happen is if what we learn has a finite basis of terms and rules, but the rules themselves allow for arbitrarily complex ‘composition’.

---

<sup>18</sup>I think it not unfair to characterize much of Fodor's argumentation to be along these lines. He is famous for remarks like “Compositionality is, as they say in Britain, non-negotiable” and “So non-negotiable is compositionality that I am not even going to tell you what it is”. Of course, there are other places where he does put forward some more serious considerations.

<sup>19</sup>Many of these are discussed in Pelletier (1994a), to which the reader is directed.

<sup>20</sup>(Fodor, 1998, p. 98) “Compositionality is a sharp sword and cutteth many knots”.

As I see it, there are three threads interwoven in the arguments:

- (i) that language is something special (infinite, or novel, or creative, or whatever),
- (ii) that people manage to use/learn/understand language despite their being “finite”,
- (iii) that one (the only?) way to do this is if language exhibited a compositional framework.

Clearly Descartes was especially impressed by (i). But he did not follow the above argumentation, and (apparently) concluded instead that people are somehow “not finite”, unlike other animals. Leibniz and others around him (see Wilson, 1989, pp. 28–32) envisaged constructing a “universal language” that could be learned very easily (because it exhibited a compositional structure?). But (apparently) these thinkers did not see natural languages as having the same properties, for, natural languages were not easy to learn and they did not represent the composition of reality in the language.

The simultaneous bringing together of (i) and (ii) occurs in the opening paragraph of Frege (1923), and can also be found in an earlier draft, Frege (1914). There is also a strong hint that he believes (iii) is the explanation of it all. Russell might also have had such a view. In Russell (1918, pp.193–195) there is recognition of the fact that (iii) would/does explain (i), but it is not clear whether this is a recommendation for a perfect language or a description of natural language. And others of the same time had similar thoughts, such as Wittgenstein (*Tractatus* 4.027–4.03) and Schlick (1918). None of these authors cite any earlier thinkers for this general argument/observation; so, although the thought seems to have been very much in the air at the time, perhaps it is true that Frege (1923) is the first time that (iii) was used as a unifying explanation. I’ve not seen any earlier ones.<sup>21</sup> Related to the Infinity of Language is also the consideration that is often called “systematicity”, perhaps first raised in the context of an argument against connectionism in Fodor and Pylyshyn (1988). This consideration is usually put: “if a speaker can understand ‘Cats chase rats’ they can also understand ‘Rats chase cats.’” Some part of this consideration seems plausible, although I think that part is already contained in one or another of the Infinity of Language arguments. But the failures of this argument in a number of other cases make me think that whatever remains after the part covered by the earlier arguments is best forgotten. (See Johnson, 2004; Pullum and Scholz, 2007, for comments.)

### 3.4.1 Holism Pro

I emphasized earlier that holistic semantic theories tended to the view that holism starts in the lexicon and that a holistic language (taken as a set of sentences) is a byproduct of this lexical phenomenon. And I said that this was in part due to the difficulties in specifying what a whole might be in the case of “all the possible meanings there are”. But there can

---

<sup>21</sup>Wilfrid Hodges likes to point to medieval Arabic scholars as precursors, see Hodges (2006). And Brendan Gillon likes to point to Pāṇini as an earlier example, see Gillon (2007). Hodges also gives (Husserl, 1900, Lecture 4) as a source, and has speculated that it is maybe possible that Frege had got his formulation from here.

be some sense where one talks of a whole as a group of the sentence-meanings. The natural examples of this are *theories* — a set of meanings/sentences that are taken to be true of, or constitute an understanding of, some area of knowledge. (For example, scientific theories.) Note that we do *not* have here a set of “all possible meanings”, but rather a set of meanings that are presumed to be well-behaved with respect to one another. (For example, are not contradictory.) In this sort of case we can talk of a different type of holism:

**Definition 5.** *Confirmation Holism is the view that a hypothesis within a theory cannot be confirmed (or disconfirmed) in isolation from the remainder of the theory.*

Note that this is a picture of how an entire theory, viewed as a collection of statements, works — as opposed to anyone’s grasp of this theory. Is it holism in the sense(s) previously discussed? Does the meaning of any sentence in the theory “depend on the meaning of all the other sentences of the theory”?

In a standard picture of a (scientific) theory, a hypothesis together with other auxiliary statements (perhaps some other hypotheses, perhaps statements about the immediate surroundings) can “generate a prediction”, which sometimes can be evaluated against the facts of the world. Confirmation holism, as characterized in Definition 5, points out that the “prediction” is in fact “generated” by the entire theory and thus it is this entire theory that is being evaluated against the facts, and not just the hypothesis that has been singled out. The point is usually put like this: experience confronts the theory as a whole; recalcitrant experiences can be accommodated either by denying the hypothesis that appeared to generate its negation or by suitably changing other parts of the theory.<sup>22</sup>

Note that, as stated, confirmation holism is about the *truth* (or truth-in-the-theory), not about meaning. It would require yet another principle that brings the two notions together, to yield meaning holism. With confirmation holism we arguably have an ontological entity — a theory— and a claim about the “direction of explanation” which is the criterion for admission into the theory. Can we view language in the same way? This seems less clear, since the criteria for admission to a language is not at all so well-defined as confirmation in a theory. But maybe it is possible to talk about how an entire society employs language, and this will be the “whole” against which the meanings of individual sentences vie for membership? This seems a possible viewpoint, but it requires a considerable amount more of background work to make it clear than has hitherto been done.

Once we have the picture of an independent-of-people theory, and what confirmation holism would look like, we can analogously consider the similar “individualistic” picture:

**Definition 6.** *Doxastic Holism is the view that one should believe what coheres with the majority of that which the person believes. The acceptance or rejection of a possible belief is due to the entire set of beliefs.*

Once again, this becomes an account of what *should* be in a theory — i.e., what *should* be believed. It takes yet further principles to turn this into semantic holism. As with

---

<sup>22</sup>Confirmation holism is usually attributed to Duhem (1906); Hempel (1950); Quine (1951).

confirmation holism, we then test sentences for possible membership into the believed set. The two types of holism are quite similar, as we can see. Confirmation holism presumes an independently-existing body of theory and evidence; doxastic holism presumes an individual with a set of believed propositions.

It should be made clear that confirmation holism and doxastic holism are not kinds of semantic holism<sup>23</sup>; for one thing, they rely on the idea that the “whole” is circumscribed in one way or another (e.g., by consistency) so that it gives a criterion for some meanings *not* being in this whole. And so they are not semantic holism by pretty much anyone’s definition. (Well, they are sometimes still called semantic holism, but only when that notion is broadened or disjunctified so as to include this other notion, as in Heal (1994). Heal requires that if one has a belief, then there are many other beliefs the person must have. But she denies that these beliefs have to be the usually-proffered sort of epistemic liaisons with items in the first belief. These other beliefs need not be further elucidations of the meaning of any terms in the first belief. It’s just that there needs to be some further beliefs.) Although confirmation and doxastic holism are not semantic holism, some have thought that semantic holism can be generated from epistemic holism by means of some further premise(s). The most common suggestion for such a premise is (some form or other of) verificationism.<sup>24</sup>

The most influential semantic theory that gives a motivation for semantic holism is inferential role semantics. So influential is this theory that it is often just taken to *be* semantic holism.

**Definition 7.** *Inferential Role Semantics is the view that the semantic content of terms is given by their liaisons with other terms.*

There are two relevant ways to understand the “liaisons” that are mentioned in Definition 7. The first, less common way is to think of them as being the connections given by antecedently-given meanings, inferences, and truths. For example, many people hold that the meaning of ‘bovine’ entails having hollow horns<sup>25</sup>, and that this is so independently of any particular person’s beliefs/knowledge of this fact about the meanings of zoological terms. Such people believe in a distinction between this type of fact and such facts as that the most common bovine in North America is the domestic cow. The former type

---

<sup>23</sup>Even the opponents of semantic holism (e.g., Fodor and Lepore, 1992, p.37) seem happy enough to embrace pure confirmation and doxastic holism.

<sup>24</sup>Quine (1951) seemed to hold that semantic holism followed from confirmation holism plus verificationism, and Fodor and Lepore (1992, pp. 216-217) also cite Harman (1974), Churchland (1984), Churchland (1986) as others who think this is right. Fodor and Lepore (1992) deny that confirmation holism plus verificationism will lead to semantic holism. On the other hand, however, Harrell (1996) and Okasha (2000) deny that Quine’s version would generate any conclusion that falls afoul of Fodor/Lepore’s problems with semantic holism. We will not delve into this dispute, referring the reader to the papers cited here.

<sup>25</sup>Rather: being an individual member of a ruminant species whose typical horned members have hollow horns.

of fact (they say) is about meanings, the latter is empirical. If all liaisons are given by inferences/truths of the meaning sort, then we might claim to have a non-individualistic meaning holism. It's a holism in an intuitive sense: the meaning of a term consists in these meaning liaisons. But of course, it also presumes these meaning liaisons to be given independently, and therefore seems to be a type of meaning atomism.

But of course, the distinction between meaning facts and empirical facts is widely thought to have been conclusively refuted by Quine (1951). And so the liaisons mentioned in Definition 7 are mostly thought to be those generated by an individual, and not to have any distinction between the person's "necessary" meaning liaisons and his "empirical" meaning liaisons — rather, they *all* form parts of the meaning of a lexical item. This is the more classical semantic holism — which is individualistic in nature. The meaning of any lexical item is the set of these liaisons — but as we saw earlier, this entails that the lexical items that thereby occur in this meaning will also in turn have this word's meaning as a part. Additionally, these lexical items in the liaisons will have liaisons with yet different lexical items, and so these further meanings too would be a part of the meaning of the first one. Ultimately, it is often said, the meaning of all lexical items is a part of the meaning of any lexical item. And so "meaning forms a unified whole" (as the holist would put it) or "the entire theory is circular" (as a compositionalist might say). As I indicated above, these liaisons would persist when lexical items are combined into larger syntactic groups, and so it would seem that every sentence would bring with it all the liaisons that any word in it contained. But since the lexical items in these liaisons have many other liaisons, these further liaisons will also be a part of the meaning of any sentence. Or to put it more broadly, the meaning of any sentence contains the meanings of all sentences.

This individualistic holism is supported by two "master arguments". The first is a plausible account of how children learn word meanings: it is a gradual process and children incorporate small parts of the meaning as how the terms interact with what terms they already have some grasp of. They learn that broccoli is food, but that it is different in shape from asparagus, but that they are both vegetables and different from chicken; that it makes Daddy happy when the child eats it, that it can come with cheese on top and still be broccoli; and so on. Plausibly, all these liaisons, and innumerable others, form a part of the meaning of 'broccoli' for the child language-learner. And what reason is there to say that they do not remain in adult meanings?

The second master argument is a slippery-slope-like argument from the apparent truism that part of the meaning of, say, 'gold', is that it is a metal. But that's not all there is to the meaning of 'gold': gold is also either yellow or white; gold is different from silver and from helium; gold is valuable; gold is what pirates bury on Caribbean islands; gold has started many wars; . . . . It seems that one cannot stop, once you get started. (Well, you might stop if you had some way to distinguish "important meaning-relations" from mere "accidental associations", along the lines mentioned earlier in a non-individualistic inferential role semantics. But here, as there, the Quinean denial of an analytic-synthetic distinction (Quine, 1951) is brought forward to assert that there is no such difference.)

Given these as the only positive arguments for meaning holism, holistic theorists have been put on the defensive, especially since the publication of Fodor and Lepore (1992) (and earlier works of Fodor's) that attacked these two considerations. Currently, straightforward arguments in favor of holism have become rare as holists instead attempt to rebut the arguments raised in Fodor's works. It seems that the attitude is that meaning holism is the correct default position, and all that is needed for establishing it is to answer the relevant pesky considerations raised in Fodor and Lepore (1992). In this regard meaning holism joins Whorfianism as "the hypothesis that just won't die" — don't give positive arguments, just rebut the contrary arguments and allow the underlying correctness of the view to shine through.

### 3.4.2 Holism Con

Most who advocate holism think of the individualistic version, where it is the lexical meanings within each individual that rely for their meaning on their liaisons with other lexical items within that individual. As I remarked above, this is different from the version of holism I was suggesting as forming the "real" conflict with compositionality, and which I attributed to Saussure. In this version, it is the public lexicon *of the language in the abstract* that has this property. The sorts of considerations that are brought against holism need to distinguish these two variants, since most of them do not hold with equal force against the two.

There seem to be two reasons that holists have gravitated toward the individualistic version. The first, rather a non-academic reason, is that it allows each person "to have their own meanings." It just is a matter of fact that many holists are drawn to some sort of individualistic relativism, and they think that the individualistic holism is a manifestation or vindication of this. The phrase 'That's true for you' can be cashed out as 'Your mental lexicon admits of a way to allow this to be true' (while retaining the idea that in my mental lexicon there is no such way). I don't propose to dwell on this feature, but I do think that anyone could walk around to the bastions of holism — many wings in philosophy departments, and "almost everybody in AI and cognitive psychology; and . . . absolutely everybody who writes literary criticism in French", as it is put in Fodor and Lepore, 1992, p. 7 (they could have added English departments and Humanities departments) — and ask why some member believes holism, this will be one of the answers.<sup>26</sup>

The second reason is due to the picture of child language learning put forward above. The incremental nature of learning that seems to accompany parental correction, especially with contrast ("No Jimmy, that's not a ball, it's called a pyramid — you see, it doesn't roll on the floor like a ball does, it just sits there") gives a reason to think that all this becomes part of the child's mental lexicon. And thus, all the liaisons form a part of the meaning of each mental item.

---

<sup>26</sup>On this particular issue, one might compare what is said in Fodor and Lepore (1992, p. xiii).

Against the individualistic version of holism there have been raised many arguments. What seems amazing to non-holists is that there should be any holists left, after they have had these arguments presented to them. As I said a few paragraphs ago, this maybe is to be explained on “non-cognitive grounds”: holism is just the right view for a (post-) postmodern humanist, and no “cognitive” consideration will ever dislodge it. (Of course, there are many who keep to holism for other reasons, and they try to show that the arguments are not as strong as they seem. Furthermore, these are arguments against *individualistic* holism, and not against the “public” version, and it is often not appreciated by non-holists that there is this other route to holism.) In this final section I present the standard arguments against individualistic holism but will not give holists the opportunity to respond. I note those that do not hold against the non-individualistic version, and will then present some considerations against that version also . . . and again will not allow its proponents to answer.

The list of anti-holism arguments directed against the individualistic version is large indeed. It is first noted that the theory predicts it to be extraordinarily likely that different people will have had different language-learning experiences, and therefore according to individualistic holism will have different liaisons among their lexical items. In turn, this means they do not speak the same language as one another. It furthermore follows that they do not mean the same thing as one another even when they utter the same word, phrase, or sentence. But then, when they (think they) agree with one another, or disagree with one another, they are in fact not doing so. Any evidence one has that you disagree with your conversational partner is, according to the theory, better evidence that you are talking about different things, and not disagreeing at all.

It is tempting to say, and many individualistic holists have, that communication does not require “absolute identity” of the corresponding mental lexicons but only that they be “similar enough”. But this can’t really be made out in a non-circular way, it seems to me. The hypothesis that two different minds are “similar enough” with respect to their understanding of some term — say, ‘democracy’ — has no empirical content other than the belief that the two people are understanding what each other says when they talk, despite their differences in acquisition of the relevant concepts. But that was precisely the (alleged) fact that the “similarity” was hypothesized to explain. The very nature of individualistic holism makes it impossible to give an independent criterion of “similar enough”.

It is also often pointed out that, not only is communication with others an impossibility according to the theory’s own features, but also one cannot communicate with oneself from one year to the next, one day to the next, one hour to the next, . . . . For, in the intervening time period the person would have new experiences and (except in very special cases where the person is comatose) these will impact the class of liaisons that impinge on the meanings of his lexicon. Although one *thinks* one remembers that the bark of eucalyptus trees peels off yearly, there is no justification for this, according to the theory. Over the year when you last had that thought, many of your liaisons have been altered — or at least, it is most likely to have happened — and thus your last-year-thoughts have no necessary connection

with this-year-thoughts.

Indeed, an individualistic holist can't change his mind about anything! For, the very act of getting new information makes it be a different thing, and hence it is *not* a case of changing one's mind about concept X. The old concept X is no longer there to have a new opinion about. Not only is it impossible to disagree with another person, as I remarked above, but it is impossible to disagree with the past. We think that we have learned that Anaxagoras was wrong when he said that the unevenness of the moon's surface is due to the mixture of earthy matter with cold. But individualistic holism is committed to claiming that we are in fact not at all disagreeing with him. We can't disagree. We can't agree either. We've just changed the subject.

The preceding arguments were directed against the individualistic or internalist version of holism. But we also considered a more "public" version, as put forward by (our version of) Saussure. In this sort of holism, it is not an individual's internal epistemic liaisons that determine the meaning of a term, but rather these liaisons are features of some external reality. However, such a position seems to diminish the first of the arguments in favor of holism — those considerations based on how an individual child's language learning is based on corrections with contrasts. For, those corrections are by their nature unique to individuals, and can't realistically be mapped onto any "external reality". Indeed, one might wonder what such an external reality could possibly be. We've seen that Saussure seems to think of it as "the French lexicon in the abstract". Could it be some sort of social entity — the language endorsed by the people in my community, or some similar definition, such as a Form of Life?

Although this version of holism seems to lose the support from the intuitive picture of language-learning, it gains some support because it avoids the criticisms based on individuals' having different liaisons. For, in this picture of meaning, the liaisons are an *external-to-individual-people* matter of fact. E.g., In the French lexicon is the fact that *mouton* is both farmed and petted as well as eaten, *redouter*, *craindre* and *avoir peur* have some specified value in terms of their contrasts with each other. This is not affected by how an individual learns these words. Plurality in French, Saussure says, has a different value from plurality in Sanskrit because unlike French, Sanskrit has a third category, dual (Saussure, 1916, p. 114). In this version of holism, meaning is not contained in, nor constrained by, how any individual learns. The object that is learned exists independently.

Notwithstanding this advantage, there still seem to be problems with this version of holism. For one thing, it is not clear how language change can be accommodated. Any attempt to allow for it seems to make the theory become susceptible to the same sort of objection we leveled against the individualistic version of holism: if the liaisons change, so that the values of the words change, then what we are saying now differs from what was said before the change — even if the same words are being used.

It is also very difficult to see how language learning works on this picture. As children, we all get pieces of this independently-existing structure; and presumably our further education gets our internal representations closer and closer to the external original. But

our learning process is fraught with individual variation — just as the individualistic holists say. So, when is it that we actually learn the language? How can we learn this antecedently-existing “lexicon of French” with all its inbuilt liaisons? And what relation does it have to our internal lexicons?

It seems to me that externalist versions of holism need to articulate a clearer notion of what the external reality is — that is, of what ontological status “the lexicon of a language” has. And they need to further explicate the relation between what an individual learns and this external item. Once this is done, there needs to be a further discussion concerning whether the external item actually plays any role in holism, or whether this externalist holism merely transfers its claims to the learned internal representations.

## 4 Leftovers, and Conclusion

One main message of this survey is that there is a fundamental cleavage both within the “compositionality” camp and the “holist” camp between an ontological version of their doctrine and a property-oriented version. And, almost all the arguments that holists have mounted against the compositionists actually are directed against the ontological version and have a very much reduced or even non-existent applicability to the property-based version. Yet, it is the property-oriented version that is of interest in philosophy of language and linguistic semantics.

Another point worthy of notice is that most versions of semantic holism are individualistic in nature. But this in turn brings up a host of problems concerning communication, belief change, understanding, and the like. I featured a non-individualistic version of semantic holism, drawing on Saussure. But this doctrine seems to have problems in formulating a way that such a language could be acquired. Generally speaking, both types of holism run afoul of the considerations that follow from the Infinity of Language.

Finally, because of my focus on versions of holism and compositionality that are opposed to one another, I have not discussed attempts to bring them together. For further thoughts along this line, one should consult Pagin (1997, 2006).

## Acknowledgments

I am quite grateful for comments on an earlier version of this paper from Chris Gauker, Peter Pagin, and Markus Werning. I wish I could have more adequately dealt with all of them. I also acknowledge comments and discussions over the years on the general topic of compositionality with Brendan Gillon, Mike Harnish, Jim Higginbotham, Wilfrid Hodges, Ali Kazmi, Ernie Lepore, Bernard Linsky, Peter Pagin, Barbara Partee, Len Schubert, Markus Werning, and Dag Westerståhl. I also acknowledge (Canadian) NSERC grant 5525 for its support over the years.

## References

- Baker, G. and P. Hacker (1980). *Wittgenstein: Understanding and Meaning*. Oxford: Oxford UP.
- Boodin, J. (1939). *The Social Mind: Foundations of Social Philosophy*. New York: Macmillan Co.
- Casati, R. and A. Varzi (1994). *Holes and Other Superficialities*. Cambridge: MIT Press.
- Churchland, P. M. (1984). *Matter and Consciousness*. Cambridge: MIT Press.
- Churchland, P. S. (1986). *Neurophilosophy*. Cambridge: MIT Press.
- Currie, G. (1982). *Frege: An Introduction to his Philosophy*. Totowa, NJ: Barnes & Noble.
- Duhem, P. (1906). *The Aim and Structure of Physical Theory*. Princeton: Princeton UP. 1954 translation of *La théorie physique, son objet et sa structure*, by P. Wiener.
- Dummett, M. (1981a). *Frege: Philosophy of Language, 2<sup>nd</sup> Edition*. Cambridge: Harvard UP.
- Dummett, M. (1981b). *The Interpretation of Frege's Philosophy*. London: Duckworth.
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Cambridge: MIT Press.
- Fodor, J. and E. Lepore (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.
- Fodor, J. and Z. Pylyshyn (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3–71.
- Frege, G. (1884). *The Foundations of Arithmetic*. Oxford: Blackwell. Translation by John Austin, 1950.
- Frege, G. (1914). Logic in mathematics. In H. Hermes, F. Kambartel, and F. Kaulbach (Eds.), *Gottlob Frege: Posthumous Writings*, pp. 72–84. Chicago: University of Chicago Press. Translated by P. Long & R. White.
- Frege, G. (1923). Compound thoughts. *Mind* 72, 1–17. Translated by R. Stoothoff; published in *Mind*.
- Garfinkel, A. (1981). *Forms of Explanation*. New Haven: Yale UP.
- Gillon, B. (2007). Pāṇini's *Aṣṭādhyāyī* and linguistic theory. *Journal of Indian Philosophy* 35, 445–468.

- Haaparanta, L. (1985). *Frege's Doctrine of Being*. Helsinki: *Acta Philosophica Fennica* **39**.
- Harman, G. (1974). Meaning and semantics. In M. Munitz and P. Unger (Eds.), *Semantics and Philosophy*, pp. 1–16. New York: SUNY Press.
- Harrell, M. (1996). Confirmation holism and semantic holism. *Synthèse* 109, 63–101.
- Heal, J. (1994). Semantic holism: Still a good buy. *Proceedings of the Aristotelian Society* 94, 325–339.
- Hempel, C. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie* 4, 41–63.
- Hintikka, J. (1980). On the any-thesis and the methodology of linguistics. *Linguistics and Philosophy* 4, 101–122.
- Hintikka, J. (1984). A hundred years later: The rise and fall of Frege's influence in language theory. *Synthèse* 59, 27–49.
- Hodges, W. (2006). Two doors to open. In D. Gabbay, S. Goncharov, and M. Zakharyashev (Eds.), *Mathematical Problems from Applied Logic I: New Logics for the 21<sup>st</sup> Century*, pp. 277–316. New York: Springer.
- Husserl, E. (1900). *Logische Untersuchungen*, BAND II. Berlin: Halle. Section references to the translation (of the 1913 2<sup>nd</sup> Edition) by J.M. Findlay, 1970, as *Logical Investigations*, Vol. 2, Routledge and Kegan Paul, London.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge: MIT Press.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford UP.
- Johnson, K. (2004). On the systematicity of language and thought. *Journal of Philosophy* 101, 111–139.
- Katz, J. and J. Fodor (1963). The structure of a semantic theory. *Language* 39, 170–210.
- Langacker, R. (1987). *Foundations of Cognitive Grammar: Vol. 1 Theoretical Prerequisites*. Stanford: Stanford UP.
- Larson, R. and G. Segal (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge: MIT Press.
- Lewis, D. and S. Lewis (1970). Holes. *Australasian Journal of Philosophy* 48, 206–212.

- Lewis, D. and S. Lewis (1996). Review of Casati and Varzi (1994). *Philosophical Review* 105, 77–79.
- Mates, B. (1950). Synonymity. *University of California Publications in Philosophy* 25, 201–226. Reprinted in L. Linsky (ed.) *Semantics and the Philosophy of Language* (Univ. Illinois Press: Urbana) 1952.
- McLaughlin, B. (1992). The rise and fall of **british** emergentism. In A. Beckermann, H. Flohr, and J. Kim (Eds.), *Emergence or Reduction?*, pp. 49–93. Berlin: Walter de Gruyter.
- O'Connor, T. and H. Y. Wong (2006). Emergent properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Okasha, S. (2000). Holism about meaning and about evidence: In defence of W. V. Quine. *Erkenntnis* 52, 39–61.
- Osgood, C., G. Suci, and P. Tannenbaum (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Pagin, P. (1997). Is compositionality compatible with holism? *Mind and Language* 12, 11–33.
- Pagin, P. (2006). Meaning holism. In E. Lepore and B. Smith (Eds.), *The Oxford Handbook of Philosophy of Language*, pp. 213–231. Oxford: Oxford UP.
- Partee, B. (1984). Compositionality. In F. Landman and F. Veltman (Eds.), *Varieties of Formal Semantics*, pp. 281–312. Dordrecht: Foris.
- Pelletier, F. J. (1994a). The principle of semantic compositionality. *Topoi* 13, 11–24. Reprinted, with additions, in S. Davis & B. Gillon (2004) *Semantics: A Reader*, Oxford UP, Oxford; pp. 133–156.
- Pelletier, F. J. (1994b). Semantic compositionality: The argument from synonymy. In R. Casati, B. Smith, and G. White (Eds.), *Philosophy and the Cognitive Sciences*, pp. 283–295. Vienna: Hölder-Pichler-Tempsky.
- Pelletier, F. J. (2000a). Did Frege believe Frege's principle? *Journal of Logic, Language, and Information* 10, 87–114.
- Pelletier, F. J. (2000b). Semantic compositionality: Free algebras and the argument from ambiguity. In M. Feller, S. Kaufmann, and M. Pauly (Eds.), *Formalizing the Dynamics of Information*, pp. 207–218. Stanford: CSLI Press.
- Phillips, D. (1976). *Holistic Thought in Social Science*. Stanford: Stanford UP.

- Pullum, G. and B. Scholz (2007). Systematicity and natural language syntax. *Croatian Journal of Philosophy* 7, 375–402.
- Putnam, H. (1954). Synonymity, and the analysis of belief sentences. *Analysis* 14, 114–122.
- Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review* 60, 20–43.
- Russell, B. (1918). Lectures on logical atomism. In R. Marsh (Ed.), *Logic and Knowledge: Essays 1901–1950*. London: Allen & Unwin. Marsh collection published 1956; page references to those in Marsh.
- Saussure, F. d. (1916). *Cours de linguistique générale*. Translated by Roy Harris as *F. de Saussure: Course in General Linguistics* (Chicago: Open Court) 1986. Page references to this translation.
- Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3, 532–631.
- Schlick, M. (1918). *General Theory of Knowledge*. New York: Springer-Verlag. Translated by A.E. Blumberg; published 1974.
- Schurz, G. (2005). Semantic holism and (non-)compositionality in scientific theories. In M. Werning, E. Machery, and G. Schurz (Eds.), *The Compositionality of Meaning and Content. Vol. 1: Foundational Issues*, pp. 271–284. Frankfurt: Ontos Verlag.
- Smuts, J. (1926). *Holism and Evolution*. London: MacMillan.
- Talmy, L. (2000). *Toward a Cognitive Semantics, Vols. I and II*. Cambridge: MIT Press.
- Westerståhl, D. (1998). On mathematical proofs of the vacuity of compositionality. *Linguistics and Philosophy* 21, 635–643.
- Westerståhl, D. (2004). On the compositional extension problem. *Linguistics and Philosophy* 33, 549–582.
- Wierzbicka, A. (1996). *Semantics: Primes and Universals*. NY: Oxford UP.
- Wilson, C. (1989). *Leibniz's Metaphysics*. Princeton: Princeton UP.