The Unbelievable Legacy of Methodological Dualism

Draft – Comments Gratefully Received

Kent Johnson
Dept. of Logic and Philosophy of Science
University of California, Irvine
johnsonk@uci.edu
http://www.lps.uci.edu/~johnsonk

ABSTRACT. Methodological dualism in linguistics occurs when its theories are subjected to general methodological standards that are inappropriate for the more developed sciences. Despite Chomsky's criticisms, methodological dualism abounds in contemporary thinking. In this paper, I treat linguistics as a scientific activity and explore some instances of dualism. By extracting some ubiquitous aspects of scientific methodology from its typically quantitative expression, I show that two recent instances of methodologically dualistic critiques of linguistics are ill-founded. I then show that the Chomskian endorsement of the scientific status of linguistic methodology is also incorrect, reflecting yet a third instance of methodological dualism.

Introduction

Perhaps more than any other discipline, linguistics has continually defended its methods and practices as "scientific". This practice has been heavily inspired by Noam Chomsky's frequent and vigorous critiques of "methodological dualism". Methodological dualism occurs in linguistics when its theories are subjected to certain general methodological standards which themselves are inappropriate for other, more developed sciences. In this vein, Chomsky has repeatedly charged many central figures in philosophy – Dummett, Davidson, Kripke, Putnam, and Quine, to name a few – of subjecting theories of language (and mind) to dualistic standards (e.g., Chomsky 1986, ch. 4; 2000, ch. 2 – 6; 1975, ch. 4). Moreover, he has argued that these standards have no known plausible defense, and that there is no reason to take them seriously. In place of these dualistic requirements, Chomsky has recommended that linguistic theories be held to the standards that normally apply to empirical theories. In his words, he advocates "an approach to the mind that considers language and similar phenomena to be elements of the natural world, to be studied by ordinary methods of empirical inquiry" (Chomsky 2000, 106). This is a natural position for Chomsky, given his rather traditional view of the relationship between philosophy and science:

> In discussing the intellectual tradition in which I believe contemporary work [sc. on language] finds its natural place, I do not make a sharp distinction between philosophy and science. The distinction, justifiable or not, is a fairly recent one…. What we call [Descartes'] "philosophical work" is not separable from his "scientific work" but is rather a component of it concerned with the conceptual foundations of science and the outer reaches of scientific speculation and (in his eyes) inference. (Chomsky 1988, 2).

If philosophy is a kind of study into the foundations of the sciences, then there is little room for a "philosophical" theory of language or mind that is not itself a "scientific" theory.

The view that linguistic theories are ordinary scientific theories, subject to the same methodological standards as the (other) sciences, has much to recommend itself. Indeed, this general view has been endorsed, at least in name, by virtually all linguists and a great many philosophers. In this paper, I adopt this view, and explore some of its consequences. I diverge from Chomsky and the many others who have discussed "scientific linguistics" in one crucial respect. It is standard to discuss the nature of scientific methodology by appealing to general descriptions of individual cases in the history of science (for a representative example, cf. the discussion of Newton in Chomsky 2000, ch. 5). Instead of considering particular examples, I focus instead on some ubiquitous quantitative aspects of ordinary scientific practice. Ordinary science encodes most of its methodology in the quantitative details of its theories, so this is a natural place to look for clues about how to do linguistics scientifically. The particular mathematical aspects I'll discuss are extremely common in ordinary scientific research, particularly when the research involves the kinds of highly complex systems that are (as we'll see) natural points of comparison with linguistics. Thus, these mathematical aspects represent some very fundamental aspects of scientific inquiry generally. Indeed, rather than applying to just a handful of case studies or one or two scientific disciplines, they are much more naturally (and normally, at least in the sciences) thought of as reflecting the actual nature of scientific inquiry quite generally.

Dialectically speaking, it won't be enough to merely point out that a number of central aspects of ordinary scientific methodology have precise parallels in linguistics. Once we take the time to extract some methodology from its quantitative presentation, it might seem obvious that the very same methods are legitimately used in linguistics. We need a foil, someone who would deny these claims. Unfortunately, such foils are all too easy to find; methodological dualism is alive and well in philosophy. My strategy, then, will be to discuss two instances of methodological dualism that have been explicitly endorsed in the literature by prominent philosophers (Francois Recanati and Jerry Fodor in particular). Abstractly stated, the two instances of dualism sound oddly familiar, which may explain why they have been adopted. The first one says that empirical theories of some phenomenon of interest must explain or account for that phenomenon. The second one says that when a theory uses theoretical terms, they must be defined. Neither principle is plausible. In debunking these principles, my interest is not so much to say something surprising (I would've taken the problems with such principles to be obvious, if they didn't keep getting endorsed). Instead, my primary aim is to expose just how deeply these principles conflict with ordinary science. As these two examples suggest, it seems that when philosophers impose dualistic standards, they don't just strike out at some controversial fringe aspect of ordinary inquiry; they go for the methodological jugular vein of scientific inquiry.

In both cases, my discussion of the principle in question takes the following form. (1) I first briefly present a proposed bit of linguistic theorizing, followed by (2) the methodological criticism to which this theorizing has been subjected. In both cases, the criticism has an air of generality about it, as though it is an instance of a much more broadly applicable methodological principle. But I then show that (3) (the generalization of) the principle conflicts with some of the absolutely fundamental and ever-present mathematical techniques of modern science, viz. certain basic statistical methods. (These techniques are used everywhere, from sociology and economics to neuroscience and biology to chemistry and physics.) Finally, I return to linguistics and show that (4) its methodology is not relevantly different from that of the (other) sciences, so that the principle in question does not apply to language.

Thus, the first two sections of this paper are a defense of current linguistic theory. In the third section, though, I bring out some divergences between linguistic and scientific methodology. Although a detailed discussion of these differences is a topic for another paper (Johnson, ms.), I offer a brief sketch of a few of these differences, showing how they constitute a third instance of methodological dualism. It is here that we precisely identify some errors of Chomsky's characterization of linguistics. I conclude in section four.

1 Dualism I: Saving the Phenomena

The first form of methodological dualism I discuss concerns the explanatory scope of a semantic theory. Francois Recanati and others have suggested that semantic theories must capture a great deal of the apparent semantic phenomena, in a sense to be spelled out below. This requirement is used to critique various types of semantic theories, such as the one developed by Herman Cappelen and Ernie Lepore (hereafter CL). I briefly sketch CL's view, and then consider Recanati's criticism. We'll then be in a position to uncover the dualistic nature behind the requirement that semantic theories "save the phenomena".

1.1 The Proposed Linguistic Theory

Recently, CL have defended a view called "Semantic Minimalism" (CL 2005). According to Semantic Minimalism, only a handful of expressions are actually context sensitive (e.g., *I, you, she, this, that, tomorrow,* etc.). There are no hidden (i.e., unpronounced, unwritten) context-sensitive elements in the syntactic or semantic structure of an expression. Thus, Semantic Minimalism contrasts sharply with most semantic theories. In particular, it is normal for semanticists and philosophers of language to assume that a correct semantics for (1a) and (2a) assigns more structure than what is given by the overt structure of these sentences.

(1) a. Mary is ready;
b. Mary is ready *to X.*

(2) a. It is raining;
b. It is raining *in location X.*

(1a) and (2a) are normally assumed to have semantic forms like (1b) and (2b). Thus, (1a) means something like Mary is ready to do something, or is ready for something to happen. Similarly, (2a) means it's raining in some contextually specified place. Semantic Minimalism denies this, holding instead that (1a) simply means that Mary is ready, and (2a) simply means that it's raining.

1.2 Criticism of the Theory via Methodological Principle

Unsurprisingly, Semantic Minimalism has encountered numerous objections (cf. CL 2005, ch. 11 – 12 for discussion). My focus will be on just one of these, which I will call the *Problem of*

Unsaved Phenomena (PUP). PUP is most straightforwardly presented in Recanati 2001 (cf. also Carsten 2004 for similar sentiments , and CL 2005, ch. 12 and citations therein). For instance, Recanati writes:

> That minimal notion of what is said is an abstraction with no psychological reality, because of the holistic nature of speaker's meaning. From a psychological point of view, we cannot separate those aspects of speaker's meaning which fills gaps in the representation associated with the sentence as a result of purely semantic interpretation, and those aspects of speaker's meaning which are optional and enrich or otherwise modify the representation in question. They are indissociable, mutually dependent aspects of a single process of pragmatic interpretation. (Recanati 2001, 88)

Recanati's pessimism about Minimalist semantic theories is driven largely by his view that such theories don't explain enough of the phenomena. In Recanati's view, a semantic theory must capture the (entire) "content of the statement as the participants in the conversation themselves would gloss it" (Recanati 2001, 79-80). Recanati expresses this in his

> "Availability Principle", according to which "what is said" must be analyzed in conformity to the intuitions shared by those who fully understand the utterance – typically the speaker and the hearer, in a normal conversational setting. This in turn supports the claim that the optional elements…(e.g., the reference to a particular time in "I've had breakfast") are indeed constitutive of what is said, despite their optional character. For if we subtract those elements, the resulting proposition no longer corresponds to the intuitive truth conditions of the utterance. (Recanati 2001, 80)

Let's begin by putting this argument in a bit more manageable form. We can characterize the Availability Principle as:

> (AP) A semantic theory is acceptable only if it correctly characterizes the intuitive truth conditions often enough within some psychologically interesting range of cases.[1]

Moreover, we can give PUP the following form:

---

[1] For present purposes, I will assume that (AP) is an appropriate formulation of Recanati's Availability Principle; any divergences between the two will not matter in this paper. One might strengthen (AP) further by specifying the particular range of cases in which a semantic theory must get things right, and by specifying how often the theory must get things right. I won't worry about such strengthenings, though; since what I have to say will apply equally to all such versions of (AP).

The Problem of Unsaved Phenomena

(3i) In all relevant ranges of cases, the intuitive truth conditions of our utterances contain much more content than what is characterized by minimalist theories.

(3ii) If (3i) is right, then from a psychological point of view, we cannot separate the minimalist aspects of meaning from those aspects supplied by a more enriched view of meaning (often enough, in any relevant range of cases).

(3iii) Hence, minimalist aspects of meaning cannot be separated from those aspects supplied by a more enriched view of meaning (often enough, in any relevant range of cases).

(3iv) But if we can't separate minimalist from non-minimalist elements of meaning (often enough, in any relevant range of cases), then minimalist theories are unacceptable.

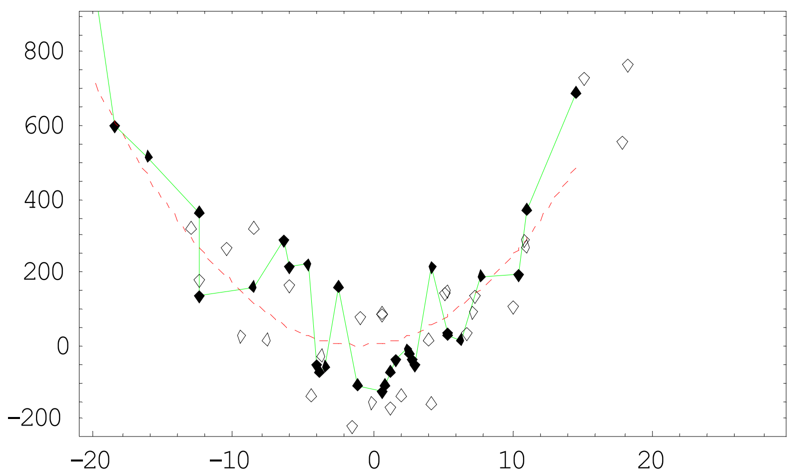
(3v) Hence, minimalist views are unacceptable.

Premise (3i) is an empirical claim; premises (3ii) and (3iv) are theoretical. Premise (3ii) comes from the quote of Recanati above (2001, 88), and premise (3iv) comes from (AP). (To see this, notice that if we can never separate out the minimalist aspects of meaning, then there must always be some non-minimalist aspects present, so the minimalist aspects of meaning never characterize the intuitive truth conditions in the utterance. Hence, by (AP), minimalist theories are unacceptable.) I won't discuss CL's attempt to deny the empirical claim (3i); suffice it to say that I find it inconclusive. But there are also flaws with (AP), (3ii) and (3iv), assuming linguistic theories are treated like ordinary scientific theories.

1.3 Is the Principle Justified by General Scientific Considerations?

In order to see what is wrong with (AP), (3ii), and (3iv), it will be useful to step back from linguistic theorizing and examine some aspects of the methodology of the (other) sciences. I'll argue that no non-linguistic scientific theories would ever be constrained to observe appropriate counterparts of these principles. The parallel between linguistic and other scientific theories thus renders (AP), (3ii) and (3iv) unacceptable.

To get things started, let's take a simple example. Suppose we are studying the relationship between different quantities of a given additive X used in some manufacturing process and the amount of some type of atmospheric pollution Y generated by the process. The industry standard is to use *n* units of X per ton of product, but for a period of time, certain companies used more or less than *n* units. The relation between the varying amounts of X used and Y emitted are given as black diamonds in the plot below (ignore the two curves and white diamonds for the moment).[2]



(The example of a pollution study here is arbitrary; it could be replaced with literally thousands of different examples from any given area of empirical science that rationally scales its measurements.) Given this data, there are (infinitely) many possible relations that could hold between X and Y. One extreme option would be to insist that every aspect of the data is crucial to understanding how X and Y are related. In such a case, a researcher might look for a function that captured the data precisely, as in the very complex one depicted with a solid line. In the present case, a polynomial of order 29, will do so, for the given raw data set of size 30:

(4) Predicted value of $Y_i = Y_i = f_2(X_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_{29} X_i^{29}$

---

[2] Zero on the x-axis represents the use of *n* units of X; other values represent the respective deviations from this standard

The resulting theory will then perfectly predict the behavior of Y on the basis of the behavior of X. The raw data, in the form of a set of pairs of measurements $\{<X_i, Y_i> : i \in I\}$, is fully accounted for. In other words, (4) saves all the phenomena, which in this case is the variation in <X, Y> scores of individual samples.

Despite its success at capturing the data, the first approach is almost never adopted. A vastly more common strategy hypothesizes a simpler relation between X and Y, and that Y is influenced by other factors unrelated to X. One might, e.g., hypothesize that that relationship is given by the simple function:

(5) Predicted value of $Y_i = f_1(X_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$

for some fixed numbers β_0, β_1, β_2. Once these numbers are determined from the data, we get the simpler curve given by the dashed line. In the present example, the values of β_0, β_1, and β_2 were determined by seeking those values for which $\sum_{i \in I}[(Y_i - f_1(X_i))^2]$ is as small as possible.

Although (5) doesn't predict the behavior of the original data as well as its rival (4), many other theoretical considerations speak in its favor. For example, suppose we got hold of another sample of data, given by the white diamonds above. Then we might ask how well the two functions captured this new data. One way to do this would be to compare the sizes of the discrepancies between what (5) and (4) predict about the value of Y for given values of X in the new data set. E.g., we might examine the ratio:

(6) $$\frac{\sum_{i \in I'}[(Y_i - f_1(X_i))^2]}{\sum_{i \in I'}[(Y_i - f_2(X_i))^2]}$$

Here *I'* indexes the second set of measurements, and f_1 and f_2 are assumed to have had the particular numerical values of their parameters – $\{\beta_0, \beta_1, \beta_2\}$ in the case of f_1, and $\{\beta_k: 0 \leq k \leq 29\}$ in the case of f_2 – fixed by the first data set. In this case, we get a value greater than 6×10^{31}, indicating that there is vastly more discrepancy between the new data and what f_1 predicts than there is between this data and what f_2 predicts.[3] Thus, the extra structure in the curve given by f_2 errs in that it captures much variance in the data that is unrelated to the true relation between X and Y.[4] In short, a bizarre model like (4) that captures all the (original) data is vastly inferior to the far more standard model like (5) that doesn't. (As a bit of terminology, I will use *model* and *theory* interchangeably.) In particular, the simpler model does a massively better job at predicting the general trends of new data as it arrives.

What then is the relation between the model in (5) and the actual raw data? This relation is given by adding a "residual" or "error" term to our equation:

$$\text{(7)} \qquad Y_i = f_1(X_i) + \varepsilon_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i,$$

The term ε_i, whose value varies as *i* varies, expresses whatever deviation is present between the model and the raw data. As (7) shows, $\varepsilon_i = Y_i - f_1(X_i)$. In practice, scientific models of complex phenomena *never* perfectly fit the data, and there is *always* a residual element (ε_i) present. This is so even when the system under study is completely deterministic. E.g., the true model might be something like

$$\text{(8)} \qquad Y_i = f_1(X_i) + f_3(Z_{1i}, \ldots, Z_{ki})$$

In such a case, Y is always an exact function of X and $Z_1, \ldots, Z_k$. However, the influence of the Z_js may be very small, very complicated, unknown, poorly understood, etc. Thus, for any number of reasons, it may be natural to model the phenomena as in (5), all the while realizing that the existence of residuals in the raw data show that there is more to the full story than is

---

[3] From a God's-eye view, this is unsurprising, because f_1 is the form that actually generated the data. I used the formula $Y_i = 3 + 4X_i + 2X_i^2 + \varepsilon_i$, where ε and X were normally distributed with a mean of zero and variances of 100 and 10 respectively.

[4] There's much more to be said about the general issues of model construction and model selection; cf. e.g. Forster and Sober 1994, Burnham and Anderson 2002 for further relevant discussion.

presented by (5). (In fact, residuals may correspond roughly to the philosophical notion of a "ceteris paribus" clause.[5])

There's nothing more basic to statistical research than the idea that the best (or true) theories/models will imperfectly fit the actual data. Indeed, that's why statistical research is founded upon probability theory instead of directly on algebra and analysis. This is just a fancy way of saying that in real empirical research of any complexity, there will always be unsaved phenomena. But this not a criticism of statistical modeling. Rather, it is a reflex of the fact that actual data is frequently the result of multiple influences, only some of which are relevant for a given project.

1.4 Is Linguistics Relevantly Different from the Other Sciences?

Let's get back to semantic theorizing. Notice that like (4) – (5), semantic theories are theories of a complex phenomenon (i.e., the interpretation of language). The raw data of a sample of the linguistic phenomena aren't numerical; instead, they are assessments about certain types of idealized[6] linguistic behavior: what sorts of things would typical speakers communicate by uttering a given sentence, and under what conditions? That is, the raw data of semantic theorizing are the intuitive truth conditions of our utterances, as we do or would make them in various contexts. Proceeding like the statistical researcher, the Semantic Minimalist begins by hypothesizing that there is some relatively simple structure – i.e. simple in comparison to the complexity of the raw data – that accounts for much of the collective behavior of the raw data. In order to obtain this simple, general structure, some aspects of the raw data (i.e., the intuitive truth

[5] The correspondence may not be perfect, though, since in real life as well as in the mathematical assumptions underlying this part of statistical modeling, the probability that the residual contributes nothing to the equation is zero. Thus, it may not be true that ceteris paribus, $Y_i = f_1(X_i)$, depending on what one's theory of ceteris paribus clauses is.

[6] The notion of idealization in linguistics and the other sciences has been discussed at great length in many places (e.g., Liu 2004, Chomsky 1986, and citations therein). Since the primary data of interest in the present paper concerns "intuitive truth conditions", the idealizations at play here are substantially less (although by no means absent!) than in other areas of linguistics.

conditions) must be ignored, just as we ignore some aspects of variance in the statistical case. In general, both minimalist semantic theories and scientific models have the same general form:[7]

(9) Raw Data = (i) *Effects* of processes under study (ii) *Interacting* in some way with (iii) *Residual* Effects

The minimalist theory supplies some aspects of meaning that are hypothesized to capture much of the general behavior of the *totality* of the data set. By assumption, the outputs of this theory are not assumed to capture all of the raw data (i.e., intuitive truth conditions of utterances). In fact, it is not even assumed that the semantic theory will *ever* capture all of the intuitive truth conditions. As we've seen, such an outcome is absolutely standard science. Our pollution researcher, would not assume that there will be some raw datum Y_i such that $Y_i = f_l(X_i)$, with no contribution from the residual effects. Indeed, it is quite typical to expect that ε_i will never equal 0, particularly when the phenomenon under study is extremely complex. (When the phenomena are quite complex, a model may be considered significant even if it captures as little as 16% of the raw data (e.g. R. Putnam 2000, 487).) Likewise, the intuitive truth conditions of utterances may always be determined by both the minimalist theory of meaning, and by other interacting aspects of communication. These other aspects of communication are familiar: background beliefs, indexical-fixing elements, demonstrations, "performance" capacities of speaker/hearers, etc.

In short, a minimalist semantic theory is just that – a *theory* about the nature of the raw data. Like any other scientific theory, one of its essential rights and obligations is to characterize those parts of the raw data it considers to be truly part of the phenomenon under study, and what other parts are due to extraneous processes; cf. (i) and (iii) in (9). The fact that semantic theories get to characterize their own scope also means that they should be judged by the standard, complicated but familiar, criteria of successful scientific theories: simplicity, elegance, predictive fecundity, integration with other successful theories which collectively account for the raw data (or, more typically, hopefully someday will account for the raw data), etc. Methodologically

---

[7] Of course, correlation does not imply causation, so more is needed here than just the regression analysis. Similar issues apply to linguistic theories as well. For simplicity's sake, I will ignore these matters, and assume that both types of models support the scientific interpretation in (9).

speaking, demanding that a semantic theory sometimes exactly characterize the intuitive truth conditions of utterances appears to be just like demanding that statistical models should (at least for some interesting range of values) be like the complex f_2, instead of the like the more standard f_1. Such a demand would be bizarre and deeply incorrect in the statistical case; I submit it is no better motivated in the case of semantic theorizing.

The points just made show that (AP) and (3iv) place an unwarranted constraint on theory construction. In no other study of complex phenomena would one demand that theories perfectly capture the raw data across some interesting range of cases. (3ii) should be rejected because it is one of the rights of a theory to provide a theoretically useful characterization of the phenomena it addresses. (3ii) simply denies this in when the theory is semantic. Thus, PUP is unsound.

Another way to view the problem with PUP is that it depends on an equivocal interpretation of "separability". Everyone can agree that the intuitive truth conditions of our utterances are almost always substantially influenced by pragmatic factors. In this sense, it's probably true that pure semantic content is "inseparable" from pragmatic factors: in actual language use, you rarely if ever find the former alone, without the latter. This interpretation of inseparability makes (3ii) plausible, but it also undermines (AP) and (3iv). After all, it's no criticism of a theory that it treats the raw data as being a product of multiple sources. If this is what separability is, then claiming that minimalist and non-minimalist aspects of meaning are inseparable simply begs the question against minimalist theories.

On the other hand, (AP) and (3iv) are plausible if inseparability means that no reasonable total theory of language will treat minimalist and non-minimalist aspects of meaning as effects of (relevantly) distinct processes. That is, in order for (AP) and (3iv) to be plausible, the relevant notion of inseparability must require that that all aspects of the intuitive truth conditions be explained by the same mechanisms. Now (AP) and (3iv) are virtually tautologies, but (3ii) loses its support. Why should the fact that the intuitive truth conditions of our utterances *do* contain both minimalist and non-minimalist aspects of content be sufficient to license the restriction that any theory of semantic content *must* capture all of these aspects? Such a view clearly begs the question against minimalist semantic theories.

I complete this section by briefly considering an argument in favor of (AP). Recanati writes:

> Suppose I am right and most sentences, perhaps all, are semantically indeterminate. What follows? That there is no such thing as 'what the sentence says' (in the standard sense in which that phrase is generally used). . . . If that is right, then we cannot sever the link between what is said and the speaker's publicly recognizable intentions. We cannot consider that something has been said, if the speech participants themselves, though they understand the utterance, are not aware that that has been said. This means that we must accept the Availability Principle. (Recanati 2001, 87 – 88)

Recanati's claim that most or all sentences are "semantically indeterminate" amounts to the claim that minimalist semantic theories don't capture the intuitive truth conditions of most or all sentences. It's unclear, though, why such a claim should be taken to imply that "there is no such thing as 'what the sentence says' (in the standard sense in which that phrase is generally used)". I take it that "what the sentence says" here refers to the content that a (minimalist or other standard) semantic theory ascribes to a sentence. If that is correct, then claiming that there "is no such thing" is simply false. Again, it's part of the job of a theory to carve out the sub-portion of the phenomenon that it directly deals with, leaving the remaining parts for further theorizing. Recanati's claim that there is no such thing as "what is said" in this context is like saying there is no such thing as the true population model f_1 in the statistical case. Hence, Recanati hasn't made a viable case for (AP).

In sum, PUP fails primarily because it ignores a basic fact about scientific theorizing: each theory gets to determine what part of the phenomena it addresses, and typically this is only a very proper subpart of the total phenomena. The requirement that a theory accommodate all of the intuitive truth conditions often enough in some relevant range of cases is a restriction on semantic theories that has no precedent in any of the developed sciences. Indeed, it is far more typical to assume that a given theory will not account for the data.

2 Dualism II: Defining Theoretical Terms

I now turn to a second form of methodological dualism. This one concerns the need to provide definitions of the technical or theoretical terms that one uses in linguistic theories. Jerry Fodor (inter alia) has stridently argued that linguists' failure to do this undermines their ability to appeal to such notions in their theories. I quickly sketch an example of the offending linguistic theory, and then we'll look more carefully at Fodor's position. We'll then consider whether it's generally true, as Fodor explicitly claims, that this requirement on theoretical terms holds generally throughout the sciences. It doesn't. Examining why the requirement doesn't hold in the (other) sciences explains why it doesn't hold in linguistic theorizing, either.

2.1 The Proposed Linguistic Theory

Within the linguistics community, it is almost universally held that there are a small number of *thematic roles* which, for present purposes, we may regard as linguistically primitive semantic elements (e.g., Baker 1988; Dowty 1991; Grimshaw 1990; Hale and Keyser 1986, 1987, 1992, 1993, 1999, 2003; Jackendoff 1983, 1987, 1990, 1997, 2002; Levin and Rappaport Hovav 1995; Parsons 1990; Pesetsky 1995). For example, one such thematic role is that of "Agency*". The thematic notion of Agency* is roughly and intuitively similar to the ordinary notion of agency, but the two are not identical.[8] In general, thematic roles are similar but not identical to certain ordinary notions. Hence, I use asterisks in the names of thematic roles to distinguish them from their ordinary language counterparts. Although Agency* isn't agency, agency can be a rough indicator of Agency*. In fact, the Agent* of a clause can be thought of as *roughly* something along the lines of the doer of the action described by the verb, if there is such a doer. (Thus, Agency* is always relativized to a sentence; Bob is the Agent* of *Bob bought a camera from Sue*, but Sue is the Agent* of *Sue sold a camera to Bob*, even though these two sentences describe the same event.)

Agency* figures into a wide variety of linguistic generalizations and explanations (cf. the citations above). For a very simple example, notice that the doer of an action is always in the subject position of a transitive verb (assuming that the verb requires either the subject or object to perform the action):

(10) a. John kicked the horse.
b. Susan kissed the bartender.
c. Christine built the shelf.

---

[8] E.g., *the poison* is plausibly the Agent* in a sentence like *The poison killed the Pope*, even though poison is not an agent of a killing.

A standard view in linguistics is that this generalization is not accidental, and in fact holds robustly across all human languages. Thus, (11) is normally taken as an important structural generalization about human languages that linguistic theories should respect and explain:

(11) All verbs with Agents* associate the Agents* with their subject positions.

There is much more to be said about (various theories of) Agency* and other thematic roles and thematic role-like elements. However, this brief introduction will be enough to introduce the methodological criticism I want to explore.

2.2 Criticism of the Theory Via Methodological Principle

A number of philosophers have taken linguists to task regarding their use of notions like Agency*. I noted above that Agency* isn't agency, so some critics have argued that linguists need to define what is meant by this new notion of Agency*. To give this worry a name, I will call it the *Problem of Undefined Terms* (PUT). The most vocal advocate of PUT is Jerry Fodor. For instance, Fodor writes,

> If a physicist explains some phenomenon by saying 'blah, blah, blah, because it was a proton…' *being a word that means proton* is not a property his explanation appeals to (though, of course, *being a proton* is). That, basically, is *why* it is not part of the physicist's responsibility to provide a linguistic theory (e.g. a semantics) for 'proton'. But the intentional sciences are different. When a psychologist says 'blah, blah, blah, because the child represents the snail as an agent…', the property of *being an Agent*-representation* (viz. *being a symbol that means Agent**) is appealed to in the explanation, and <u>the psychologist owes an account of what property that is</u>. The physicist is responsible for *being a proton* but not for *being a proton-concept*; the psychologist is responsible for *being an Agent*-concept* but not for *being an Agent*-concept-ascription*. <u>Both the physicist and the psychologist is required to theorize about the properties he ascribes</u>, and neither is required to theorize about the properties of the language he uses to ascribe them. The difference is that the psychologist is working one level up. (Fodor 1998, 59, underlining added; cf. Fodor and Lepore 2005, 353 – 4 for similar sentiments)

Prima facie, Fodor's argument here seems pretty compelling.[9] If *Agency** doesn't mean agency, then what does it mean? Demanding that one define one's terms seems like a reasonable and

---

[9] It is somewhat odd that Fodor endorses PUT. After all, the linguist posits the notion of an Agent* as a linguistically primitive element, and Fodor has vigorously defended the view that such (the concepts denoted by) linguistically primitive elements can't be defined. Instead, he maintains that the best theory of what they mean is simply given atomistically. According to atomism, the best theory of meaning says only that the word *dog* means dog (or better, *dog* denotes the concept of dogs). One would think that such an attitude would carry over to other parts of language, too: *Agen*t* means Agent*.

important requirement. After all, if you don't know what Agency* means, then how can you use it in an alleged "generalization" like (11)? If you only say that the subject of transitive verbs can be the Agent* of the verb, but no independent constraints are placed on what it is to be an Agent*, then the alleged generalization about verbs has no content. To see this, just replace the technical term *Agent** with any other made-up word, say *flurg* (cf. Fodor 1998, 59). Now (11) can be restated as *All verbs with flurgs associate the flurgs with their subject positions*. Obviously, with no theory of what flurgs are, this statement is empty.

2.3 Is the Principle Justified by General Scientific Considerations?

As his allusion to physics makes clear, Fodor believes that any scientific theory must "provide an account of what property" is denoted by any theoretical term it uses. Indeed, his criticism is just that linguistics violates this general principle of science. Alas, this principle is not generally true in the sciences. It's often the case that a theoretical term is introduced to denote some hypothetical property that is posited in the theory in order to account for some kind of "surprise" or "pattern" in the empirical data. Moreover, the nature of this theoretical property is often determined not by stipulation in advance, but by continuing theoretical work. This pattern of theory construction is especially common in the earlier, developing stages of some area of inquiry, which is undeniably where all areas of linguistics currently are at (cf. §3 for some discussion of this last claim). In that sense, it is common for a scientist to "theorize about the properties he ascribes", but this does not amount to producing more of "an account of what property" he postulates than the linguist produces regarding Agency* and the like. Furthermore, just like our first dualism, this aspect of scientific inquiry is encoded in the mathematical apparatus that constitutes our standard forms of data analysis.

To flesh out these ideas, let's consider an example. As before, I stress that this is only an example. It is meant to illustrate what scientists standardly do when exploring complex phenomena. Suppose we are examining the concentrations of three chemicals X, Y, and Z in a given region. One hundred groundwater samples are taken from the region, and the amounts of each of X, Y, and Z are recorded. When the data are plotted as points on three axes, they are distributed as in Figure 1 below. Rather than being randomly dispersed, the data appear to be more structured around a general "pattern". This structure is of course a real surprise, in the sense that it's extraordinarily improbable that a random sample of unrelated measurements would ever yield such a pattern. (The boxes are scaled to a 1-1-1 ratio to visually present the correlations, as opposed to the covariances, of the three variables.)

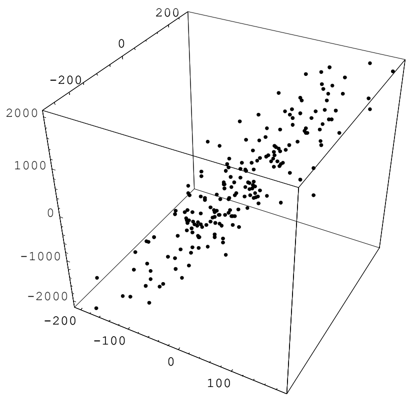


Figure 1

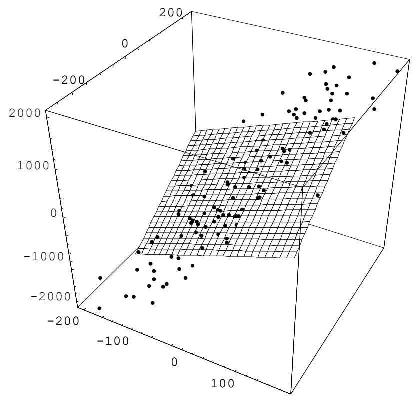


Figure 2

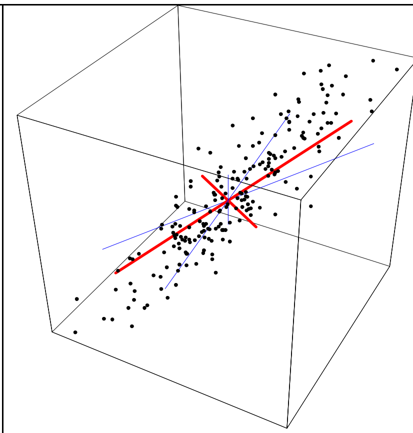


Figure 3

It's the essence of the sciences not to ignore such patterns in the world. A natural first step is to try to understand "how much" of a pattern is there, and what its nature is. Obviously, the relative concentrations of X, Y, and Z are related. From the geometric perspective of the cube, the data appear to be organized around an angled plane, partly depicted in Figure 2. The fit isn't perfect, and the planar surface lies at a skewed angle, so all three axes of the cube are involved. But if we used a *different* set of axes, we could view the data as organized primarily along two axes. In other words, suppose we replaced axes X, Y, and Z with three new axes, A, B, and C. (If we keep A, B, and C perpendicular to one another, we can think of ourselves as holding the data fixed in space, but rotating the cube.) Moreover, suppose that we choose these axes so that the A axis runs right through the center of the swarm of data. In other words, let A to be that single axis on which we find as much of the variation in the data as possible. If we had to represent all the variation in our data with just one axis, A would be our best choice. It wouldn't perfectly reproduce all the information about X, Y, and Z, but it would capture a lot of it. Suppose that we now fixed the second axis B in such a way that it captured as much of the remaining variation in the data as possible, after we factor out that variation that is represented by the A axis. Together, A and B would determine a plane lurking in the three-dimensional space. (The two darker lines in Figure 3 correspond to Axes A and B.) By projecting all the data onto this plane, we could recover quite a bit of information about the variation in the data. We won't recover all the information in the original three-dimensional space, but we'll recover a lot of it. (We'll miss just that information regarding how far to one side or another from the plane the actual data points lie.) If we set axis C to best capture the remaining information, we will then be able to recover all the information in the original space. If we represent the original data using all three new axes, we will be merely representing the original data as linear combinations of X, Y, and Z. If, however, we are satisfied with the amount of information we get using only one or two axes, we can represent the data in a less complex manner, using only one or two dimensions, instead of the original three-dimensional format. Whether we represent the pattern in the data with just axis A, or with axes A and B is not a decision that the mathematical analysis itself makes for us, although in many cases other techniques or considerations will provide strong evidence for one option.

The technique just described is called Principal Component Analysis (PCA). The mathematical essence of PCA involves finding new axes on which to represent the data. The first Principal Component (PC) is axis A, and the second and third PCs are B and C respectively. Clearly, this technique is not restricted to three dimensions – it can be used with any (finite) collection $X_1,\ldots,X_n$ of measurements.[10] The real scientific import of PCA comes when we find that e.g., one or two PCs can account for say 95% of the variation in one or two hundred types of observations. Geometrically, this is like finding that in a space of one or two hundred dimensions, all the data are arranged almost perfectly in a straight line or on a single two-dimensional plane lurking in that space. That is a pattern far too extreme to be random, and it

---

[10] Very briefly, here is a general description of the mathematics of PCA. You first construct an $n \times n$ correlation (or covariance) matrix, where the *ij*th entry gives the correlation between X_i and X_j. One then extracts all the eigenvalues (there are virtually always n of them) and eigenvectors of unit length from this matrix. Ordering these eigenvectors according to the size of their eigenvalues, we obtain our PCs. The *k*th eigenvector gives the direction of the *k*th axis in the n-dimensional data space. Moreover, the *k*th eigenvalue expresses the amount of total variation in the data that is captured by the *k*th PC. (When working with a correlation matrix, the total variation will always be n.) Also, the amount of variation in a given measurement X_i that the *k*th PC accounts for is given as $\sqrt{\lambda_k}\, a_{ki}$, where λ_k is the *k*th eigenvalue and a_{ik} is the *i*th element of the *k*th eigenvector. This form of PCA produces perpendicular axes, each of which accounts for the maximum amount of remaining variance in the data. Once one decides to retain m PCs (where m is almost always much less than n), the basis for the m-dimensional subspace can be changed to suit background hypotheses, etc.

cries out for explanation. PCA and related techniques can help quantify this pattern in useful ways.

By exposing a small number of dimensions of variation which capture most of the variation in the data, we derive an *explanandum*. Why should the concentrations of three – or more realistically, 30 or 100 – different chemicals behave as though they were from only two sources? At this point, an empirical/metaphysical hypothesis suggests itself: maybe they behave this way because there are two sources responsible for the emission of the chemicals. At this point in the investigation, we may not be able to say much about these two hypothesized sources other than that they are what are emitting the chemicals in question. We can't, for instance, automatically infer that one source corresponds to axis A and the other source to B. Axes A and B determine a plane in the data-space, but infinitely many other pairs of lines (not necessarily at right angles) could also determine that same plane. (In fact, I used the three lighter axes in Figure 3, which are not at right angles to one another, to generate the data.[11]) If the two sources do correspond to axes other than A and B, this means that they each emit different concentrations of X, Y, and Z than A and B predict. If the sources correspond to two axes that are not at right angles, this means that their emissions of chemicals are correlated with one another. Finally, even the notion of a "source" must be understood in a broad functional sense. It's possible that there are two physical sources that both emit the same relative concentrations of X, Y, and Z, and so one of the axes represents both their emissions. Alternatively, one axis could represent the joint effects of several physical sources that emit different concentrations of X, Y, and Z, but which are all perfectly correlated with one another. At the same time, the hypothesis that there are two sources of emission is a very strong and testable hypothesis, not least because together they must nearly determine the particular plane in the data-space (cf. e.g., Malinowski 2002, ch. 10 – 12 for discussion of this and other uses of PCA and related techniques in chemistry).[12]

---

[11] The two longer axes occur on the plane (given by $4X + 5Y - Z = 3$), and the third axis is just Z. The data was generated by randomly selecting points (coded on the X, Y, Z axes) on each axis from a normal distribution with mean 0 and variance 50. These points were then weighted at .65, .35, and .05 and added together to produce the data.

[12] There are *lots* of other ways to explore the results of the PCA. If two candidate sources are found, the PCA will supply evidence about their correlation. If the sources are different factories, this may indicate e.g., the degree to which they are working together, or are both influenced by the same economic factors, etc. Also, if the two sources

PCA is one member of a family of methods – which includes factor analysis, structural equation modeling, latent class analysis, discriminant analysis, multidimensional scaling, and others – for exploring the extent to which there is latent structure in the data. These techniques all involve the uncovering of underlying regularities that appear when individuals (persons, groundwater samples, etc.) are measured in a variety of different ways. (Such techniques bear some similarity to Whewell's (1840) "consilience of inductions", although they supply much more information and structure.) In the study of complex phenomena, where many different sorts of measurements are possible, the use of such techniques is extremely common, particularly in the early stages of inquiry, but also consistently throughout the development of the theory. Indeed, one would be hard-pressed to identify a line of scientific inquiry into some complex phenomena where such techniques *weren't* used. It's just what you do with data.

In short, in the early stages of ordinary scientific inquiry, it is perfectly possible to hypothesize the existence of unobserved empirical structures, sources, etc. without having much of a theory about their natures. It's certainly correct that as the investigation develops, the chemist "owes an account" of the nature of the sources of emission. However, at the early stages of inquiry, the details of this account may be a long ways off. Indeed, for highly complex situations, there may be a lengthy initial period of many rounds of data analysis, with a focus only on figuring out how many latent structures there are and what kinds of overt measurements they affect. If scientists were required to precisely characterize all hypothesized structures, a great deal of successful research into complex systems would be illegitimate.

2.4 Is Linguistics Relevantly Different from the Other Sciences?

If we consider the underlying logic of quantitative methods like PCA, here is what we find. Sometimes there are one or more significant PCs lurking in a multidimensional data-space. If both the dimensionality of the data-space and the amount of variation that the PCs account for are large, then we have a significant explanandum that needs explaining. This need can justify the provisional adoption of a *hypothesis* that there is some unobserved empirical structure underpinning these mathematical PCs. Of course, the hypothesis may be wrong, and even if it is

---

are discovered using only some of the measurements, say X and Y, then the PCA will express the relative amounts of Z that the sources emit. If Z is a noxious pollutant, this may be extremely important information.

on the right track, much of the nature of this unobserved structure is an issue for further research. Importantly, though, the multivariate nature of the PCs constrains (and thus helps to form) hypotheses about the unobserved structures responsible for the PCs. That is, the fact that the PCs are built out of multiple overt measurements severely limits what sorts of things they could represent. E.g. inspection of our simple example above shows that if you know what a proposed PC predicts on just one dimension, say X, then you can tightly constrain what it predicts on the remaining dimensions. Put another way, not all sets of possible predictions correspond to possible PCs.

The situation with linguistics is very similar to what we have just seen with quantitative data analysis. For starters, the discipline of linguistics is certainly still at an early "exploratory" stage, and it certainly concerns a very complex phenomenon. Aspects of these phenomena are represented with various measurements, which themselves aren't quantitative, but instead concern such things as the grammaticality or acceptability of a sentence, its sound and meaning, etc. Linguistic theorizing is currently centrally driven by examining various patterns that exist within various interestingly clustered sets of sentences. The goal is to uncover the latent structures responsible for (much of) these patterns. Moreover, just as with PCA, the structures we hypothesize may not capture all the empirical data – linguistic theories will employ residual effects, as we saw in §1. Thus, it's natural to understand thematic roles like Agency* as having the same sort of epistemic status as the latent structures in any other exploratory data analysis. True, we don't know fully what Agency* is, but that doesn't mean that we can't provisionally hypothesize the existence of such a latent element as part of a theory about why our overt measurements (i.e., linguistic judgments) behave as they do. In the early stages of inquiry, one chooses to provisionally hypothesize the existence of thematic roles or of a correlate of a highly significant PC for largely the same reasons: both types of hypotheses are testable in many ways and have lots of room for potentially wide-ranging augmentation and refinement through further scientific inquiry.

So why did the Problem of Undefined Terms seem so compelling? I suspect that there are two main reasons for this.[13] First, PUT encourages us to think of a *completed* theory of Agency*.

---

[13] Actually, there is also probably a third reason, which has to do with the unclarity of some of the crucial judgments, and the fact that such unclarity often seems to accumulate as theories become increasingly complex.

In a finished linguistics, we would expect more details about Agency*. But the real issue is about justifying the very *beginnings* of such a theory. Why it is worthwhile to explore a theory that posits thematic roles rather than some other one, say one that posits tiny fairies in our brains?

Second, recall that the real worry behind PUT was that a technical notion like Agency* is too unconstrained and underdetermined to be useful in theorizing or in forming generalizations. If there are no constraints on what it is to be an Agent*, then one can make a generalization like *All (or only) Agents* are Fs* true by brute force. For any potential counterexample C to the generalization, nothing prevents you from stipulating that C is not an Agent*. But this worry is defused when we notice that, just as PCs can only be (non-trivially) extracted from collections of more than one sort of measurement (e.g. X, Y, and Z concentrations), the extraction of linguistic structure always involves multiple sorts of linguistic phenomena. To see this, consider the following simplified sketch of how one might justify positing Agency* in a linguistic theory. Since I only want to illustrate a very common method in linguistics, I'll omit lots of details and data, in order to avoid the complexities of doing linguistics straight out. You don't have to be convinced of the details of the example in order to understand the method employed. (All the ideas and data, though, are very familiar from the linguistics literature.)

Suppose a linguist, call her Lana, notices that some derived nominals (nouns that Lana hypothesizes to be derived from an underlying verb) have a form that corresponds to the passive form of the verb (12), but other verbs do not (13):

(12) a. Sharon proved the theorem/ the theorem's proof by Sharon
b. John destroyed the vase/ the vase's destruction by John
c. John created the vase/ the vase's creation by John

(13) a. Sue loved Mary/ *Mary's love by Sue
b. Sue resembled Mary/ *Mary's resemblance by Sue
c. Sue awakened / *the awakening by Sue

There is a lot to be said about this issue. However, since it does not pertain directly to the Problem of Undefined terms, I will leave this topic for another day.

After studying this pattern, Lana begins to explore the hypothesis that there is some structural property she calls *flurg* present in the nouns or verbs in (12) and absent in (13), or vice versa. At this point, flurg simply encodes a difference between two kinds of words. But then Lana notices that with nominals like (12), the nominal can be the complement of a possessive, or it can have a passive *by*-phrase adjunct, but not both (although it can take some *by*-phrases and possessives):

(14) a. The Roman's destruction of the city
b. The destruction of the city by the barbarians
c. *The Roman's destruction of the city by the barbarians
d. The Roman's destruction of the city by catapults and mass attack

Lana now hypothesizes that such nominals have some property that can license either the possessive or the passive *by*-phrase, but not both. She then hypothesizes, that this property is flurg, the same one used in (12) – (13). As a third bit of data, Lana notices that languages lack symmetric pairs of verbs for asymmetric events. E.g., while we have verbs like *kick*, *lift, build*, etc., we don't have verbs like *blik*, where x bliks y if and only if y kicks x (and similarly for *lift, build,* etc.). (In contrast, notice that we do find symmetric pairs elsewhere; e.g., the direct and indirect objects of *Sue sent a letter to Tim* and *Sue sent Tim a letter* can be exchanged without any apparent change in meaning.) Now Lana hypothesizes that flurg is once again responsible for this phenomenon, because flurg must necessarily be located in the subject position of the verb, and can never appear in object position. Of course, any of these hypotheses could turn out to be false, but so far Lana feels that her developing theory of flurg and its roles in language is sufficiently plausible to merit further study.

As Lana continues to study the words that she hypothesizes have flurg versus those that don't, she begins to sense that words with flurg all share a certain semantic similarity, although she cannot fully articulate what it is. To explore this hunch, she gives a very brief characterization of what little she knows about this semantic similarity to a variety of people (both linguistically trained and untrained). She uses only a couple words as examples, and then gives her subjects a large number of other words, and asks them to indicate whether they perceive this hypothetical semantic property in the words, and if so, where (e.g., in subject or object position). To her surprise, there is an enormous amount of agreement across subjects.

They typically find flurg clearly present or absent in the same places, and are unsure about the same cases. As Grice and Strawson (1956) noted, this kind of agreement marks a distinction (between the presence and absence of flurg), even if many details about the nature of the distinction are unknown. Attending to the cases where there are uniformly clear judgments, Lana notices that these are also places where the other hypothesized effects of flurg occur. Thus, she further expands her hypothesis about flurg, claiming that it is, or is associated with some kind of semantic feature, which she is currently investigating, but has not yet fully identified or even confirmed. Since this hypothesized semantic feature of flurg corresponds roughly but not completely to the notion of agency, she renames *flurg Agent** out of convenience.

Notice that in the story just told, the nature of flurg is unconstrained only at the very beginning, when it is hypothesized to underwrite just one distinction. However, as the theory is developed to account for multiple distinctions, flurg becomes more tightly constrained. By the end, the hypothesis that flurg exists is rather demanding. It is not enough for subjects to simply feel that flurg is clearly present/absent in a given word; the theory also makes (heavily ceteris paribus, as it turns out) predictions about the syntactic behavior of that word, and it predicts where in the word, if anywhere, subjects will sense flurg's semantic presence. The multiple dimensions that are used to characterize flurg render it anything but an unconstrained hypothetical element.

In PCA, one seeks out a few vectors in the data space that explain most of the variation in the data. In linguistics one seeks out a few structural elements that explain most of the variation in the data. In neither case are these elements always fully defined or understood. Demanding a complete definition in the linguistic case is a methodologically dualistic standard. I know of no defense of such an atypical stance towards linguistics. Certainly Fodor's incorrect appeal to the sciences (quoted above) offers no such support. As a final comment, it's worth observing that the parallel in linguistics with techniques like PCA is quite strong. Indeed, work in consensus theory suggests that it might be possible to perform some form of latent variable analysis on the kinds of collective linguistic judgments discussed above (e.g., Batchelder and Romney 1988). In such a case, the evidence for a structural element like Agency* could be reduced to something like the question of whether the first PC in the data space captured nearly all the variation in the data. This topic has not been researched by theoretical linguists.

3 Dualism III: Aggregation and Degrees of Accuracy

In the previous two sections, we considered two dualistic principles which, though rejected, attempted to make linguistic theorizing *harder* in some respects than ordinary scientific theorizing. At this point, it's natural to ask, are there any commonly accepted aspects of linguistic methodology that make linguistic theorizing *easier* than scientific theorizing? In fact, there are, and in this section, I gesture at some of them. My remarks here will be brief, although I address this issue in more detail elsewhere (Johnson, ms.).

Before beginning, two caveats are in order. First, I'll illustrate the divergences between linguistic and (other) scientific methodologies with a particular example. But as will be obvious to anyone familiar with mainstream linguistic methods, the morals of this case study generalize very broadly to a vast amount of linguistic research. Also, since the purpose of the example is only to illustrate certain widely used *methods* of linguistic theorizing, I make no effort to exhaustively characterize the relevant literature. For present purposes, that would only complicate matters by presenting a more complex instance of the same methods my simplistic example will bring out. (Indeed, the precise details are unimportant enough that readers familiar with linguistic methods may wish to skip the example altogether, and go right to the discussion.)

Second, my critical remarks are not intended to be some sort of "tearing down" of linguistic theory or the "harassing of emerging disciplines" (Chomsky 2000, 60, 77). Rather, I suggest only that linguistics should follow the quantitative sciences, which routinely study their own methodologies in order to better understand, use, and improve them.[14] Without such study, there will continue to be many reasons why Chomsky was incorrect to claim that ordinary linguistic practices have "exhausted the methods of science" (Chomsky 1986, 252). With these caveats in hand, let us turn to the example.

In a series of papers, Norbert Hornstein (1998, 1999, 2000a, 2000b, 2004) has argued that control phenomena can be accounted for simply by allowing movement into theta positions. E.g., the relevant syntax of (15a) does not have the traditional form in (15b), where PRO is a distinct lexical item controlled by *Sue*. Instead, the proper form is in (15c), where *Sue* has moved from

---

[14] To verify this last claim, one need merely consult a current statistics journal, or a journal of a particular science that publishes papers on mathematical methods (e.g., *The Review of Economics and Statistics* or *The Journal of Mathematical Psychology*).

the lower subject position to the higher one. (Following Hornstein, I treat movement as a combination of the Minimalist operations of Copy and Merge.)

(15) a. Sue wants to win;

b. Sue_i wants [PRO_i to win];

c. Sue wants ~~Sue~~ to win.

More generally, Hornstein argue that linguistic theories don't need to posit the null pronomial element PRO at all. Nor do linguistic theories need to posit a control module to determine the referent of an occurrence of PRO. Hornstein's argument is primarily driven by considerations of simplicity. PRO is unnecessary in a theory, he argues, because the phenomena that initially motivate positing PRO can be accounted for by appealing to independently motivated components of the grammar. Movement (aka Copy and Merge), Hornstein assumes, is a prevalent feature of grammar. If all the relevant facts can be accounted for without positing PRO, then (with other things being equal, as well), linguistic theories should favor the simpler theory and reject the employment of PRO. In the development of this theory, Hornstein also notes several advantages to his theory. Here are two representative examples.

It's well-known that both PRO and traces (i.e., the residue of Copy and Merge) are both phonetically null. By identifying the two, we reduce the need to explain this fact from two separate phenomena to just one. Similarly, the fact that *wanna* contraction can occur either with raising or with control will now require only one explanation:

(16) a. I seem to be getting taller.

a'. I seemta be getting taller.

b. I want to get taller.

b'. I wanna get taller.

Famously, there is some need to explain this type of contraction, because it does not happen willy-nilly. You can turn *You want to help Mary* into a *wanna*-question by asking *Who do you wanna help?*, but you can turn *You want John to help Mary* into a question only by asking *Who do you want to help Mary?* It is not grammatical to ask **Who do you wanna help Mary?*

As a second advantage, Hornstein considers "hygienic" verbs, as in *Peter washed/dressed/shaved Dan*. These verbs are interesting, because they can also appear intransitively:

(17) Peter washed/dressed/shaved

Notice that (17) has a reflexive meaning; it says that Peter washed *himself* (or shaved himself etc.) This stands in stark contrast to other verbs that can drop their objects; *John ate* does not mean John ate himself, only that he ate something. According to standard views, the reflexive behavior of (17) is puzzling, since a reflexive reading would most naturally be supplied by an element like PRO, but PRO is typically thought to appear only as the subject of a clause (e.g., Chomsky and Lasnik 1993).[15] But on Hornstein's view, the syntactic element PRO is replaced by whatever structure underlies movement phenomena, and it is well-known that movement can occur from object position – e.g., on Hornstein's view, *who did Shaun kiss ~~who~~$_i$* would be an example.[16] Hornstein shows how the reflexive readings in (17) are (relatively) neatly and unproblematically produced within his theory.

Unsurprisingly, Hornstein's proposal has not gone unnoticed (e.g., Brody 1999, Culicover and Jackendoff 2001, Landau 2000, 2003, Manzini and Roussou 2000). Here are two representative criticisms of the view. The first problem comes from Landau 2003. Landau argues that Hornstein's theory has problems accounting for *partial control*, illustrated in (18):

(18) The chair wanted to meet on Tuesday afternoon.

(18) is most naturally interpreted as expressing that the chair wanted some set X to meet on Tuesday afternoon, where X contains the chair and at least one other person. Roughly and intuitively speaking, partial control constructions are distinctive in that the controlling DP is only a proper subset of the collective subject of the embedded clause. Further evidence that there is a

---

[15] For instance, *Bill wants Mary to kiss* cannot have the structure *Bill$_i$ wants Mary$_j$ to kiss PRO$_{i/j}$*; it can mean neither that Bill wants Mary to kiss Bill nor that Bill wants Mary to kiss herself.

[16] On a technical note, it is standard to distinguish (roughly speaking) A-movement from A'-movement. However, if one relinquishes, as Hornstein does, the constraint that a syntactically realized NP (or DP, I won't adjudicate here) must bear at most one theta role, such a distinction becomes less well motivated. As Hornstein discusses at length, a central component of his theory is that syntactic chains can bear multiple theta roles. This relaxation of the Theta Criterion in many ways is a central bit of machinery of his theory.

plural syntactic subject in the lower clause comes from the ability of partial control to support distinctively plural types of predicates and anaphors:

(19) a. Susan enjoyed getting together on weekends.
b. Steve wondered whether helping one another would be productive in the long run.

It is hard to see how a "control as movement" view such as Hornstein's can handle partial control. The relevant structure of (18) is simply *The chair*$_i$ *wanted to* ~~*the chair*~~$_i$ *meet on Tuesday afternoon*. Nothing in Hornstein's theory appears to explain how the overt copy of *the chair* denotes a single person, but the deleted copy of that very expression denotes a group, of which the chair is only one member. This suggests that the subject of the lower clause of (18) is realized as something other than merely a copy of *the chair*. Further evidence that this other element may be PRO comes from the fact that partial control does not appear to exist in raising constructions:

(20) *The chair seemed to meet on Tuesday afternoon.

"Without further detail", Landau argues, "one can already see how damaging the very existence of partial control is to the thesis 'control is raising'. Simply put: *there is no partial raising*. It is not even clear how to formulate a rule of NP-movement that would yield a chain with nonidentical copies" (Landau 2003, 493).

The second problem comes from Brody (1999, 218 – 19). Consider the following pattern.

(21) a. John attempted to leave.
b. *John was attempted to leave.
c. *John believed to have left.
d. John was believed to have left.

Why can't (21c) be used to express that John believed that he himself had left, just as (21a) expresses (roughly) that John attempted to make himself leave? Similarly, why can't (21b) be used to express that someone attempted to make John leave, just like (21d) expresses that someone believed that John had left? If control is just movement, as Hornstein proposes, then we have no explanation for the different syntactic abilities of what is typically thought to be NP-movement – (21c,d) – and what is typically thought to be control – (21a,b).

Now, of course there is a great deal more to be said about Hornstein's theory. The theory has more prospects and problems than what I've presented, and there are objections and replies to them, and objections and replies to the objections and replies, and so on. But the points to follow can be made by examining just a few considerations.

In ordinary scientific inquiry, there are a great many questions regarding the relation between the empirical data, background assumptions, and the resulting theory or theories generated from them. Neither the questions nor the answers are inherently quantitative, although with statistical modeling, they are typically expressed that way. As linguistics is currently practiced, though, there is no known way to address the vast majority of these questions. Of the many questions that are routinely studied in the other sciences but not linguistics, one is an 800-pound gorilla. I call it the *problem of model-performance aggregation,* or the *problem of aggregation* for short. Intuitively speaking, the problem is that current linguistic methods don't provide any systematic means for aggregating multiple assessments of a theory into a single more general assessment. To see what I mean here, consider the following example.

Imagine a linguist who needs to evaluate Hornstein's view. Perhaps, e.g., she works in a related area of syntax or semantics, and she is trying to decide whether a movement analysis of control is promising enough that she should explore incorporating this view into her own theory. (Obviously, if she has little faith in the view, she will be disinclined to expand her own theory in this direction.) Simplifying greatly, suppose also that her only considerations about the view concern the advantages and disadvantages just listed: she thinks that (i) Hornstein's theory does a good job accounting for certain reflexive intransitives and (ii) for *wanna*-contraction. But she thinks (iii) the theory is weaker at handling partial control and (iv) passivization. Our linguist recognizes that Hornstein's theory can be made to handle (iii) – (iv), but she also thinks that the only way to do so is rather unelegant and somewhat ad hoc. (For the moment, let's bracket the very difficult issue of how she arrived at these judgments.) Now what does she do? How should she combine her various assessments about Hornstein's theory to arrive at a single assessment? At this point, linguistic methodology comes to a grinding halt. If our linguist has no further data or considerations to add, she cannot further analyze the situation except by appealing to her subjective impressions (and perhaps also the impressions of her colleagues) of the overall promise of the theory. This is the problem of aggregation: linguistic methodology provides no theoretical tools to guide the inference from the collection of considerations to an overall assessment of the theory.

Although the problem of aggregation is rather obvious, its importance shouldn't be underestimated. Its force can be brought out with some very naïve questions. Is the overall assessment of the theory in question positive or negative in light of (i) – (iv)? Does the advantage of (i) outweigh the disadvantage of (iv)? If so, by how much? Enough to offset the sum disadvantage left over from (iii) when the advantage of (ii) is factored out? If we factor in another feature, say (v) the relative simplicity of the theory, is the theory preferable to a given rival theory? What if we encounter some further data, and we are unsure of how it impacts on Hornstein's theory? Suppose our linguist is also considering a rival theory that she feels does well on (i) and (iii) but not so well on (ii) and (iv), but she also feels it does reasonably well with respect to some further area (vi). How is she to decide between them? Linguistic methodology provides no systematic means for addressing any of these questions. You just have to go with your gut.

The fact that linguistic methodology offers no way to aggregate various aspects of the performance of the model on actual empirical data stands in stark contrast to the (other) sciences. Indeed, it's fair to say that the ability to aggregate diverse aspects of the performance of a model is the single most central aspect involved in the study of the relation between a model and the empirical phenomena it models. We've already seen one simple form of assessing how well overall a given model does with respect to some data. In our discussion of (5) and (6) above, we considered the sum of squared deviations of the data from the model's predictions: $\sum_{i \in I}[(Y_i - f_1(X_i))^2]$. This quantity represents an aggregation, via addition, of all model's failings to accurately predict the data. It appears in a variety of techniques for analyzing and assessing the overall performance of a given theory. Moreover, there are a large number of sophisticated techniques that aggregate such diverse aspects of a theory as its empirical coverage and its "simplicity". For instance, much attention has been paid to (various forms of) the Akaike Information Criterion (e.g., Forster and Sober 1994, Burnham and Anderson 2002). Additionally, there are a wide variety of other methods of "model selection" (e.g., Zucchini 2000) but they all centrally involve the aggregation of various features of the model into a single assessment.

Is the problem of aggregation really a problem? For analyses that require the aggregation of different aspects of a model's performance, can't we trust the reflective, considered judgments of professional linguists? In practice, of course, that's what we do, but the present question concerns the reliability of such an analytic and inferential strategy. Although I don't know of any research on linguistics, decision analysts have paid a great deal of attention to the reliability of professional (e.g., scientific, medical, academic) judgments. The short answer is that scientists' intuitive assessments of theories are subject to the same foibles that pervade ordinary human judgment and decision-making. In particular, scientists are apt to be overconfident about the accuracy, success, and promise of their favored theories, and underconfident about their rivals' theories (e.g., Frischoff and Henrion 1986; cf. Bishop and Trout 2005 for detailed discussion and many more citations). Additionally, this lack of inferential guidance in linguistic theorizing renders it virtually *impossible* for linguists to recognize certain sorts of "surprising results" that are fairly common in the other sciences, and which we may expect to obtain in linguistics. (E.g., there are many instances where what is commonly agreed to be a good/bad theory is seen to be not so, only after the use of more systematic, principled, and precise instruments than professional judgments.) Such facts explain why the non-linguistic sciences maintain active research programs into the study and refinement of methods for inferring theories from the empirical data and background hypotheses. Indeed, a central part of applied statistics and related subject-specific areas (e.g., psychophysics, biometrics) involves developing fine-tuned methods, often restricted to very specific sorts of models and inferences, that can be used in place of human judgment. It is a commonplace that the careful use of such methods produces better – often dramatically better – results than unaided judgments.

The problem of aggregation is only one particularly striking instance of a divergence between linguistic and (other) scientific methods. In general, it is very hard to see how to systematically combine any effects or features of a linguistic theory into a single collective judgment. Even more generally, linguistic methodology provides remarkably little in the way of systematic study and analysis of the relation between one's empirical data and one's theory (although it does provide some). In contrast, such methods are a central component of the (other) sciences. To give a sense of this, I will end by briefly sketching another area where statistical methods have been highly developed for assessing one's theory in light of the data. (On a technical note, the quantitative models I have discussed have all been continuous, rationally scaled variables. All points made in this paper could just as easily be made with nominally scaled variables, which correspond to measurements such as yes/no, on/off, present/absent, grammatical/ungrammatical, answer A/B/C/D, etc.)

Statistical models and inferences rely on background assumptions, typically quite substantial ones, such as that the residuals ε_i discussed above are all normally distributed (i.e., they fall on a "bell curve"). Typically many of these background conditions are false. An important question in statistical modeling concerns how sensitive the model and its predictions are to various kinds of violations of these assumptions. If a model's predictions vary substantially with relatively minor violations of the assumptions, the model and inferences from it are said to be "fragile", or to lack "robustness". Fragility also appears when a model and its inferences are heavily influenced by "minor" aspects of the data that is selected to derive the model (and/or its parameters) But as Leamer (1985, 308) notes, "a fragile inference is not worth taking seriously." This is because the background assumptions typically are false, and models (particularly models of complex phenomena with substantial residual effects) that are highly sensitive to small changes in empirical data are unreliable.

The very same issues arise for theoretical linguistics. Mainstream linguistic theories rely heavily on a whole host of background assumptions, primarily in the form of idealizations. We make various idealizations about the speakers of the language, their judgments about various expressions, the computational architecture of grammatical competence, the nature of Universal Grammar, etc. These idealizations are important, perhaps crucial, for real-life work in linguistics. But they are probably also all false. If we accept that "a fragile inference is not worth taking seriously", we need to develop better methods for analyzing how well various linguistic theories perform even when some of the idealizations above are relaxed in various ways. Furthermore, we need to develop better methods for understanding and estimating the degree of dependence of individual theories on potentially "variable" linguistic judgments of various sorts. And in assessing various theories, we should penalize, perhaps severely, theories that are too sensitive in either of these ways. The systematic study of the sensitivity of models has played virtually no role in contemporary linguistics.

In addition to the issues discussed here, ordinary scientific research involves quite a number of other forms of analysis of the theory-data relationship which raise corresponding questions about this relationship in linguistics. A few examples include such notions as tolerance, leverage, reliability, validity, and the identifiability of a model's parameters. Although these notions raise questions about linguistic theorizing, there has been essentially no work in theoretical linguistics to address these issues. Until at least some of these issues are addressed more conscientiously, we must realize that in some important respects, linguistic theories are held to different standards than ordinary scientific theories. In short, in many respects, it is not even remotely true that ordinary linguistic methods have "exhausted the methods of science" (Chomsky 1986, 252).

4 Conclusion

If we treat linguistics as a genuine scientific activity, then such philosophical issues as the Problem of Unsaved Phenomena and the Problem of Undefined Terms simply disappear. At the same time, a wide range of new issues arise regarding the study and improvement of linguistic methods. Linguistics is rather unique among the sciences in that it has seen relatively little work on its methodology. But linguistics is arguably the area of inquiry that needs such work the most.

On a final note, one occasionally encounters the view that it's optional whether one treats linguistic theories as scientific, and that linguistic theories can alternatively be treated as "philosophical" theories. I confess I simply don't understand this position. Nonetheless, there are a few things that can be said in response. First, such alternative theories don't appear to compete or conflict with anything I've said about scientific theories. Second, I mean very little by calling a semantic theory "scientific". Linguistic theories deserve this appellate, I suggest, primarily because their construction and confirmation centrally involve employing some of our best known methods for obtaining knowledge about a particular empirical phenomenon. From this perspective, it is unclear how one could reasonably defend the importance of a "non-scientific" theory of language. (It's trivial to show that there are other sorts of projects; the trick is to interpret linguistic theorizing as some other project that is interesting and worth persuing.) Moreover, my present use of the idea that linguistic theories are a type of scientific theory is, I believe, especially uncontentious. So for an alternative view to avoid the conclusions we've drawn, one needs to show why the particular features of linguistic theorizing that I've appealed to are not part of some other (worthwhile) form of linguistics.

References

Baker, Mark 1988, *Incorporation*, Chicago: University of Chicago Press.

Basilevsky, Alexander 1994, *Statistical Factor Analysis and Related Methods*. New York: Wiley-Interscience.

Burnham, Kenneth, and David R. Anderson (2002). *Model Selection and Multimodel Inference* (2nd ed.). New York: Springer.

Cappelen, Herman, and Ernie Lepore, 2005. *Insensitive Semantics*. Oxford: Blackwell's.

Carsten, Rachael 2004, "Relevance Theory and the Saying/Implicating Distinction." In L. Horn and G. Ward (eds.) *Handbook of Pragmatics*. Oxford: Blackwell, 633 – 656.

Chomsky, Noam 2000. *New Directions in the Study of Language and Mind*. Cambridge: CUP.

Chomsky, Noam 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomksy, Noam 1988. *Language and Problems of Knowledge.* Cambridge, MA: MIT Press.

Chomksy, Noam 1975, *Reflections on Language*. New York: Pantheon.

Chomsky, Noam 1980, *Rules and Representations*. New York: Columbia University Press.

Noam Chomsky 1986, *Knowledge of Language*, Westport, Conn.: Praeger.

Chomsky, Noam, and Howard Lasnik 1993. "The Theory of Principles and Parameters". Reprinted in Chomsky 1995.

Davidson, Donald 1986, "A Nice Derangement of Epitaphs", in Lepore, Ernest (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell's.

David Dowty 1991, "Thematic Proto-Roles and Argument Selection", *Language*, Vol. 67, no. 3, pp. 547-619.

Fodor, Jerry A. 1998, *Concepts*, Oxford: Clarendon.

Fodor, Jerry A. and Ernie Lepore 2005, "Impossible Words: A Reply to Kent Johnson". *Mind and Language* 20:3, 353 – 356.

Forster, Malcom, and Elliott Sober 1994. "How to Tell when Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions" *British Journal for the Philosophy of Science* 45. 1 – 35.

Gleitman Lila R., and Mark Liberman (eds.) 1995. *An Introduction to Cognitive Science, Volume I: Language*. Cambridge, MA: MIT Press.

Grice, H. P. and P. F. Strawson 1956. "In Defense of a Dogma". *The Philosophical Review* 65:2, 141 – 158.

Jane Grimshaw 1990, *Argument Structure*, Cambridge: MIT Press.

Kenneth Hale and Samuel Jay Keyser 1986, "Some Transitivity Alternations in English", Lexicon Project Working Papers 7, Cambridge, MA: Center for Cognitive Science, MIT.

Kenneth Hale and Samuel Jay Keyser 1987, "A View from the Middle", Lexicon Project Working Papers 10, Cambridge, MA: Center for Cognitive Science, MIT.

Kenneth Hale and Samuel Jay Keyser 1992, "The Syntactic Character of Thematic Structure", in I. M. Roca (ed.) *Thematic Structure and its Role in Grammar*, New York: Foris, pp. 107-143.

Kenneth Hale and Samuel Jay Keyser 1993, "On Argument Structure and the Lexical Expression

of Syntactic Relations”, in K. Hale and S. Keyser (eds.), *The View from Building 20*, Cambridge, MA: MIT Press, pp. 53-109.
Kenneth Hale and Samuel Jay Keyser 1999, “A Response to Fodor and Lepore, ‘Impossible Words?’”, *Linguistic Inquiry*, Vol. 30, no. 3, pp. 453-66.
Hale, K. and Keyser, S. J. 2003, *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.
Norbert Hornstein 1999, “Movement and Control”, *Lingusitic Inquiry*, Vol. 30, no. 1, pp. 69-96.
Norbert Hornstein 2003, “On Control”, in Randall Hendrick (ed.) *Minimalist Syntax*. Oxford: Blackwell, 6 – 81.
Ray Jackendoff 1983, *Semantics and Cognition*, Cambridge, MA: MIT Press.
Ray Jackendoff 1987, “The Status of Thematic Relations in Linguistic Theory”, *Linguistic Inquiry*, Vol. 18, no. 3, pp. 369-411.
Ray Jackendoff 1990, *Semantic Structures*, Cambridge, MA: MIT Press.
Jackendoff, Ray 1997, *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
Jackendoff, Ray 2002, *Foundations of Language*. Oxford: OUP.
Johnson, Kent 2004, “From Impossible words to Conceptual Structure: The Role of Structure and Processes in the Lexicon” *Mind and Language* 19:3. pp 334 – 358.
Johnson, Kent ms. “Methodological Divergences between Linguistics and the (Other) Sciences”
Leamer, Edward 1985, “Sensitivity Analyses Would Help”, *American Economic Review* 75. 308 – 313.
Beth Levin and Malka Rappaport Hovav 1995, *Unaccusativity: At the Syntax--Lexical Semantics Interface*, Cambridge, MA: MIT Press.
Malinowski, Edmund R. 2002, *Factor Analysis in Chemistry*. New York: John Wiley and Sons.
Terence Parsons 1990, *Events in the Semantics of English*, Cambridge, MA: MIT Press.
Pesetsky, David, 1995. *Zero Syntax*. Cambridge, MA: MIT Press.
Pietroski, Paul 2005. *Events and Semantic Architecture.* Oxford: OUP.
Putnam, Robert D. 2000, *Bowling Alone* New York: Touchstone.
Recanati, Francois 2001. “What is Said”, *Synthese* 128. 75 – 91.
Stone, Tony, and Martin Davies 2002, “Chomsky Amongst the Philosophers”. *Mind and Language* 17:3, 276 – 289.
Whewell, William 1840, *The Philosophy of the Inductive Sciences*. London.
Zucchini, Walter 2000. “An Introduction to Model Selection”. *Journal of Mathematical Psychology* 44, 41 – 61.