

'Kriptenstein's' Skeptical Paradox and Chomsky's Reply

Barbara Scholz

University of Toledo

Abstract: Chomsky's *KNOWLEDGE OF LANGUAGE* addresses certain conceptual questions about the foundations of generative linguistics that center on a 'skeptical paradox' that Kripke attributes to Wittgenstein. Chomsky's discussion offers an extended defense of his psychological conception of grammar against this challenge. This essay argues that Chomsky's response to the skeptical paradox is inadequate, but instructively so. The inadequacies of Chomsky's reply surface as a destructive dilemma for the psycholinguist conceptually committed to the generative paradigm in such a way as to reveal a conceptual incoherence in that paradigm. Specifically, the essay exhibits the dilemma as it arises for the performance theory of Berwick and Weinberg (1986). While modification of the philosophical foundations of generative linguistics may show the working psycholinguist the way out of the dilemma, this essay leaves the dispute unresolved, making only the negative point.

1. Introduction

In *KNOWLEDGE OF LANGUAGE*, Noam Chomsky (1986) focuses attention on three questions that are fundamental to generative linguistics conceived as a branch of psychology. The questions are:

- I. What constitutes knowledge of language?
- II. How is knowledge of language acquired?
- III. How is knowledge of language put to use? (1986:3)

Questions (I) and (III) are striking insofar as providing an adequate response to each is as much a philosophical as a linguistic project and touch on traditional philosophical issues concerning the nature of mind, language, and thought. In the context of articulating his response to (III), Chomsky discusses the so called 'skeptical paradox' that Saul Kripke attributes to Ludwig Wittgenstein in *ON RULES AND PRIVATE LANGUAGE* (1984). Chomsky takes this paradox to pose a deep challenge to the philosophical foundations of psychological linguistics.

The objectives of this essay are twofold. First, it aims to make clear the foundational philosophical challenge the skeptical paradox poses for linguistics. Secondly, it attempts to evaluate the force and adequacy of Chomsky's replies to

this challenge. I shall argue that Chomsky's response does not address the central charge that there are conceptual inadequacies in linguistics conceived as a branch of psychology. While I think that the prospects for finding a solution to the paradox are reasonably good, it is not within the scope of this essay to articulate what I think is the most promising solution. I leave the dispute unresolved, having made only the negative point.

2. The Skeptical Paradox

The skeptical paradox that Kripke attributes to Wittgenstein (hereafter Kriptenstein¹) is a family of arguments in the form of a *reductio ad absurdum*. Throughout his exposition of the Wittgensteinian texts, Kripke focuses on the idea that word meaning and denotation is rule-governed. The paradox, however, is perfectly general in the sense that a member of the family of arguments applies to any behavior that is alleged to be rule-governed. The version of the paradox that will interest us is displayed below.

The Paradox

- (1) Jones knows a language L by being in competence state S.
- (2) If Jones knows L by being in competence state S, there must be some fact about Jones that constitutes his being in S that justifies claims that he is in that state.
- (3) There is no neutrally specifiable² fact about Jones that constitutes his being in S and that justifies claims that he is in that state.
- (4) Hence, it is not the case that Jones knows a language L by being in state S.
- (5) (1) and (4) are incompatible.

¹ In *ON RULES AND PRIVATE LANGUAGE*, Saul Kripke presents a forceful interpretation of arguments found in Ludwig Wittgenstein's *PHILOSOPHICAL INVESTIGATIONS* and *REMARKS ON THE FOUNDATIONS OF MATHEMATICS*. In *KNOWLEDGE OF LANGUAGE*, Chomsky's primary concern is with the argument that Kripke attributes to Wittgenstein and eschews discussion of the exegetical question whether Kripke's Wittgenstein adequately represents Wittgenstein. In this paper, we shall follow Chomsky and concern ourselves only with the force of Kripke's reading of Wittgenstein for linguistics in the generative paradigm.

² The specification of the fact that constitutes the grasp of a rule must provide non-trivial necessary and sufficient conditions that do not assume that what constitutes using a rule has already been explicated.

All versions of the paradox assume (1) and (2). (3) is established by argument. The first premise attributes a psychologically real competence state to a speaker. (2) unpacks necessary conditions for that attribution. It is assumed that the facts that constitute S also justify attribution of S to a speaker.

Given the ruling philosophical realist idea that psychological state attributions are fact stating, the skeptical paradox challenges the philosopher *cum* linguist to specify or describe the kind of thing, at a high level of generality, that constitutes such a state. The demand seems fair, if the philosopher/linguist holds that there must be facts that constitute S, it is fair to ask him to say what kind of thing those facts are.

In its paradigmatic and most analytical form, the task of philosophical semantic theories has been understood as one of providing an analysis or informative explanation of what constitutes meaning and reference. The idea is that an adequate philosophical semantic theory will answer the question: in virtue of what does a token of 'A' mean A? The question is an ontological question about speaker meaning and understanding of meaning. An adequate ontological analysis or explanation will provide a non-trivial specification or description of the facts, states of affairs or states of mind that constitute internal representation and use of rules. It is assumed that an answer to the ontological question will also specify the epistemological ground of rule attributions. Of course, such a description will not describe specific experimental effects, but it will answer the general question of what state S is in a way that allows experimental data to be interpreted to warrant the attribution of S. That there is a neutral description of these rules and their use (i.e., a description that does not somehow assume the notion of rule-following) is the idea that is the target of the *reductio*.

Kripke makes the assumption (for *reductio*) explicit as follows:

By means of my external symbolic representation and my internal mental representation, I 'grasp' the rule for addition. One point is crucial to my grasp of this rule. Although I myself have computed only finitely many sums in the past, the rule determines my answer for indefinitely many sums that I have never previously considered. This is the whole point of the notion that in learning to add I grasp a rule: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future. (1984:8)

The point of claiming that meaning and knowledge of language are constituted by the grasp of rules is to capture the idea that linguistic behavior is normative. This notion is specified by the following adequacy conditions.

Adequacy Conditions on descriptions of competence states:

Condition A: Internally represented rules determine (in some sense) future and as yet unconsidered linguistic behavior.

Condition B: These rules are uniquely³ represented.

When a philosopher claims that linguistic behavior is normative, he typically means (at least) that it is behavior of which it makes sense to claim that it is correct or incorrect. Of a normative behavior it is intelligible to say that it was mistaken or in error. Inextricably bound with the idea of a normative phenomenon is the idea of a unique standard in virtue of which that phenomenon is judged permissible or not. Of course this idea of linguistic normativity contrasts with the linguist's standard conception of the prescriptivity of grammars. Introductory linguistics textbooks take pains to deny that linguists' grammars are prescriptive and deny that grammars (not behaviors) are normative. The import is to distinguish between the grammars of scientific linguistics and old fashioned grammarians' grammars. There is no conflict between what the philosopher claims is normative and what the linguist claims is not normative, since behaviors are not grammars.

Rule-governed behavior contrasts sharply with behavior that is merely rule-conforming. The description of rule-conforming behavior need satisfy neither Adequacy Condition A nor B. For example, the behavior of bodies conforms with Newton's Laws, which are rules of a sort, but bodies do not make mistakes if their behavior does not conform with those laws. Moreover, by continuing in its orbit, Neptune merely conforms with Newton's laws and does not apply an internal representation of them. Satisfaction of Adequacy Conditions A and B captures the notion that linguistic behavior is governed by internally represented rules. What is wanted in response to the paradox is an informative explanation of what it is to internally represent and apply a rule that does not assume that this notion has been satisfactorily explicated. A neutral description shows (3) is false.

Computational theories of mind can profitably be seen as attempts to provide neutral descriptions of what it is to follow a rule. For example, in the second chapter of *THE LANGUAGE OF THOUGHT* (1975) Jerry Fodor motivates his appeal to the computer metaphor with a discussion of Wittgenstein's skeptical paradox. The idea, of course, is that a computational theory of mind can neutrally specify facts in virtue of which a speaker internally represents and uses rules. Having a "language

³ 'Unique' should not be read here to exclude the possibility of genuinely ambiguous syntactical tokens. Chomsky often states this adequacy condition by claiming that the rules must be correct. For example, he writes that the idea of an internal grammar or I-language is correct, while in the case of an E-language there is no issue of correctness or incorrectness (1986:26). Alternatively, he claims that a generative grammar "purports to depict exactly what one knows when one knows a language" (1986:24).

of thought" i.e., a language that the machine is built to use, and something like a compiler that translates from natural languages into a brain code, constitutes a neutral description of rules and their use.⁴ While Fodor's suggested solution is well worth discussion, we shall not pursue it here. Our concern is specifically with Chomsky's views in *KNOWLEDGE OF LANGUAGE* where he is not concerned with computational theories of what it is to mentally represent and use a rule (1986:239)

So far, we have said nothing about the argument for (3), the claim that there is no fact specifiable in neutral terms that constitutes the grasp of a rule so that Adequacy Conditions A and B are satisfied. The argument for (3) is an argument by elimination. Kriptenstein considers and rejects a variety of candidates which, if adequate, would show that (3) is false. In general, the candidates fall into two categories based on the way they fail as solutions to the paradox. The first group of candidates (mental images, experiential states, dispositions) fail because they do not satisfy Adequacy Conditions A and B. In consequence, such candidate solutions fail to capture the relevant properties of linguistic behavior. Kripke's arguments against the idea that linguistic behavior can be explicated in terms of a speaker's dispositions are reminiscent of Chomsky's attack on Skinner. The second group of

⁴ Too briefly, the debate with respect to whether computational or language of thought solutions to the paradox are adequate focuses on whether a causal description of mental content is adequate. For given a description of the content of mental representations and their use in purely causal (neutral terms), what Fodor calls the disjunction problem arises. In *PSYCHOSEMANTICS* (1987:102) Fodor describes the disjunction problem as follows.

We can put it that a viable causal theory of content has to acknowledge two kinds of cases where there are disjoint causally sufficient conditions for the tokenings of a symbol: the case where the content of the symbol is disjunctive ('A' expresses the property of being (A \vee B)) and the case where the content of the symbol is not disjunctive and some tokenings are false ('A' expresses the property of being A, and B-caused 'A' tokenings misrepresent). ...

The disjunction problem is extremely robust; so far as I know, it arises in one guise or another for every causal theory of content that has been thus far proposed.

The disjunction problem is the problem of distinguishing the case where 'A' correctly represents (A \vee B) and the case where 'A' misrepresents B, because it was caused by B. The problem arises because descriptions of mental representations (couched in purely causal terms) do not satisfy the adequacy conditions (much discussed in text above) on normative or rule-governed phenomena. Mental content like linguistic knowledge is both productive and unique in the relevant respect. The disjunction problem shows that a purely causal theory of content cannot capture the notion of a mistaken representation.

candidates, which includes linguists' descriptions of competence states, i.e., grammars, fails to show that (3) is false, because they do not neutrally describe internally represented rules and their use.

If my exposition of the paradox has been clear, then it should be obvious why the typical description of a competence state does not adequately neutrally describe the fact that constitutes an internally represented rule (or system of rules). What is wanted is an account of what it is for an individual to represent and apply rules in terms that make no appeal to the notion of a mentally represented rule. The explanandum must not appear in the explanans on pain of circularity. Our conception of a typical description of such competence states assumes the very idea that is to be explained -- that a speaker represents and uses rules. Hence, a competence state as typically conceived and described begs the very question at issue.

To be sure, Chomskian competence states satisfy Adequacy Conditions A and B. Such states are claimed to be explanatorily adequate, and not merely descriptively adequate and so satisfy Adequacy Condition B. Moreover, it is typically claimed that such states and their description account for the phenomena of linguistic productivity and so by their very nature satisfy Adequacy Condition A. But it is not a solution to the skeptical paradox insofar as it does not describe a fact in neutral terms that shows (3) of the paradox to be false.

Kripstein puts the point this way:

. . . our understanding of 'competence' is dependent on our understanding of 'following a rule'. . . Only after the skeptical problem about rules has been resolved can we then define 'competence' in terms of rule following. . . Although the remarks in the text warn against the use of the 'competence' notion as a solution to our problem, in no way are they arguments against the notion itself. Nevertheless, given the skeptical nature of Wittgenstein's solution to his problem . . . , it is clear that if Wittgenstein's standpoint is accepted, the notion of 'competence' will be seen in a radically different way from the way it is implicitly seen in much of the linguistics literature. For if statements attributing rule-following are neither regarded as stating facts, nor to be thought of as explaining behavior, it would seem that the use of the ideas of rules and of competence in linguistics needs serious reconsideration. (1984:31)

If Kripke is right, then if the linguist is going to resist the contradiction at (5) -- that a speaker both knows L by being in S and does not know L by being in S, then he must do so by denying (2). It is with the rejection of (2) that the paradox becomes interesting.

(2) claims that a speaker is in S only if there is a fact (specifiable in neutral terms) that constitutes that state, satisfies Adequacy Conditions A and B, and justifies attribution of S to the speaker. The second premise will be false just in case speakers represent and use rules (in some sense), but one of the following four responses to the paradox is viable.

Four Possible Responses to the Paradox:

Response A: the facts that constitute S do not satisfy Adequacy Conditions A and B, or

Response B: the facts that constitute S do not justify or warrant competence state attributions, or

Response C: the facts that constitute S cannot be specified or described in other terms, or

Response D: there are no facts at all that constitute S.

Since competence states (for Chomsky) are described by appeal to rules, the linguist must claim that (2) not (3) is false if he is to avoid the antinomy at (5). In itself, rejecting (2) is not especially problematic *prima facie*. For there are four ways the linguist can avoid the conclusion that a speaker both is and is not in S. Some of these alternatives are more attractive than others, but the antinomy is well worth avoiding.

It will be useful to attend to exactly what the linguist accepts should he embrace one of the Responses A - D. To begin, consider Response A. Accepting that the facts that constitute competence states do not satisfy Adequacy Conditions A and B entails giving up either the idea that the rules that constitute competence states are unique or that linguistic knowledge is productive. Neither alternative is acceptable. Suppose, for example, that two rules of S, R and R', when applied to a string P assign a different status to P (e.g. noise and acceptable), so the rules of S are not unique. If both R and R' are claimed to be the standard that constitutes S in virtue of which P has the status it has, then S is no standard. Non-unique standards are not standards at all. Thus, the Adequacy Conditions seem essential for psychological linguistics.

Alternatively, the linguist might consider rejecting Adequacy Condition A as a means of accepting Response A. In this case he would reject the idea that knowledge of language is productive in the sense that what people know when they know a language applies to expressions they have never heard or previously considered. This option is not open to one who claims that language is rule-governed. Accepting Response A by rejecting Adequacy Conditions A or B would require a revision in the typical conception of competence.

Alternatively, consider Response B. The Chomskian idea is that facts about the speaker's competence states (e.g. psycholinguistic evidence) justify the attribution of such states to speakers. If the linguist accepts Response B, he denies that claim. If one wants to take psycholinguistic evidence to be evidence about competence states and the use of rules, then one cannot accept Response B.

Next, consider Response C. In accepting Response C the linguist accepts that facts that constitute S cannot be described without appeal to internally represented rules. The advocate of such a view might claim that the facts in virtue of which Jones knows L by being in S are brute linguistic facts that cannot be described in other terms. Such a view suggests that there is a wide and unbridgeable gulf between linguistic facts and physical and chemical facts, for example, which are specifiable without appeal to internally represented rules. *Prima facie* it also conflicts with naturalism, the view that linguistic facts are part of the natural biological order. Indeed, the view seems to entail a form of linguistic dualism on which the facts of linguistics are a *sui generis* kind of entity -- an internal-rule-fact.

Finally, accepting Response D entails that whatever represented rules are they are not psychologically real phenomena. In consequence of Response D, competence state attributions are not literally true or false, although they may be useful for some purpose. The linguist who accepts Response D adopts a form of instrumentalism about competence states.

At this point, the Chomskian linguist is faced with four possible responses to the skeptical paradox. From what we have said so far, Response C is the most attractive alternative --but much more on this below.

3. Chomsky's Response to Kripstein

In KNOWLEDGE OF LANGUAGE, Chomsky argues that the skeptical paradox does not show that "the notion of competence [must] be seen in a light radically different from the way that it is seen in much of the linguistics literature".⁵ Chomsky's defense consists in accepting Response A and Response C but explicitly denying Response B and Response D.

⁵ In Chapter 4 of KNOWLEDGE OF LANGUAGE Chomsky focuses a great deal of attention to arguing against the skeptical solution that Kripke attributes to Wittgenstein. Wittgenstein's solution to the paradox, like Chomsky's, is skeptical insofar as it accepts that (3) is true but (2) is false. Kripke claims that the thesis that there is such a thing as a private language follows as a corollary from Wittgenstein's own skeptical solution. Whether Wittgenstein's skeptical solution is adequate and whether the impossibility of a private language (whatever that is) does so follow is independent of the challenge that the paradox raises for the generative linguist.

We have already noted that as a matter of logic, the generative linguist must accept at least one of the above disjuncts. In this section, I shall argue that it is not open to the Chomskian linguist to accept Response A and that if Response C is accepted, the problem of saying what it is to use one rule rather than another resurfaces in the project of formulating performance models of competence theories.

When Kriptenstein claims that linguistic behavior is normative, what is claimed is that the Adequacy Conditions apply. In *KNOWLEDGE OF LANGUAGE*, Chomsky explicitly argues that linguistic phenomena are not normative (as *per* Response A) and urges that all issues of correct and incorrect performance can be 'dropped' by considering 'normal'⁶ cases of attributing rules to native speakers. Chomsky illustrates the claim by observing that we do say that a child's internal rules are incorrect, but we are unlikely to say of adult (normal) native speakers that their rules are incorrect. So of children who overgeneralize and say "slepped" Chomsky writes:

we will say that their rules are "incorrect" meaning different from those of the adult community or a selected portion of it. Here we invoke the normative teleological aspect of the common sense notion of language. (1986:227)

By contrast, we do not say of the adult Irishman who says "There himself goes down the road," that his internal rules are incorrect. According to Chomsky, the generative linguist can embrace Response A, because the linguists' theory merely describes a speaker's internally represented rules.

Accepting Response A as a way to avoid the antinomy is fundamentally misguided. First, what is at issue (with respect to the paradox) is the normative status of linguistic behavior, not the normative status of the description of the competence state. For example, an anthropologist may claim to describe a system of moral rules in a particular community. While the anthropologist's description is not normative, what he describes, if accurate, will characterize morally permissible and impermissible (morally normative) behavior in that community. Similarly, the linguists' competence theory describes a competence state, nevertheless the hypothesized internally represented rules characterize the sentences of L which are linguistically (not morally) permissible or impermissible. Make no mistake, our concern is not with what one calls this interesting property of linguistic behavior. What it is important to see is that one of the primary reasons to posit internal rules that are used by speakers is to explicate what the philosopher (erroneously if you like) calls linguistic normativity. If one accepts Response A and so denies that the

⁶ One wonders what Chomsky means by 'normal' here. One suspects that he means something like 'in cases where error in performance is not at issue', but then it is trivially and uninterestingly true that issues of distinguishing correct and incorrect rule attributions do not arise in those cases.

facts that constitute internal rules and their application must satisfactorily explain linguistic productivity or provide a standard of performance, then one rejects the idea that language is a rule-governed phenomenon. Accepting Response A entails denying distinctive characteristics of the phenomenon that one wants to explain.

Secondly, Chomsky's point that linguistic competence is not a normative notion (because we would say of the child that his rules are incorrect but would not say of the Irishman that his rules are incorrect) is moot. What is relevant is that if the child overgeneralizes the use of a rule, for example, that **Verb+PAST** --> **Verb+/d/**, then it is correct for the child to say "sleeped". What is correct or incorrect is behavior relative to internally represented rules not descriptions of those rules. In consequence, Chomsky's attempts to avoid the paradox by claiming the generative linguist can coherently accept Response A is not persuasive. Indeed, if successful, his arguments would undermine his own performance/competence distinctions⁷ in the sense that a competence state provides a standard of performance and explicates linguistic productivity.

Chomsky explicitly advocates two routes out of the paradox, for he also endorses the idea that the linguist can avoid the antinomy at (5) by claiming that (2) is false in virtue of there being *sui generis* linguistic facts. I shall argue that if the generative linguist embraces this claim (Response C) the troubles of the skeptical paradox resurface, but now as a problem for the psycholinguist. In short, my thesis is that the challenge that the skeptical paradox presents for the linguist is a bump-under-the-rug phenomenon. If the linguist attempts to detoxify the paradox by claiming (2) is false in virtue of accepting Response C, then the paradox revisits itself on the working psycholinguist. But first, we need a clearer idea of what these *sui generis* rule-facts are supposed to be like.

Chomsky has recently taken to using the neologism 'I-language' to refer to what he previously called competence states. For Chomsky, an I-language is "some

⁷ Space does not permit consideration of the many versions of the performance/competence distinction as made by Chomsky in the course of his long and illustrious career. Suffice it to say that at one time Chomsky seemed to think that a competence theory was to be distinguished from a performance theory only insofar as the competence theory required an idealization away from various interfering performance factors, e.g., memory limitations, background noise, etc. It is not clear that this notion of the competence/performance theory is normative in the sense I have been concerned with here. Linguistic normativity is at best a troublesome notion. Unfortunately, it is not at all clear that either the philosopher of language or the linguist can live without it. The intuitive distinction between rule-conforming and rule-guided behavior seems cogent. If the linguists' theory does not capture the relevant features of the phenomena, then it seems that his theory simply does not explain something that needs to be explained. A full scale study of linguistic normativity in conjunction with an examination of various versions of the performance/competence distinction would be useful.

element of the mind of a person who knows a language, acquired by the learner, and used by the speaker." [1986:22] In particular, it is a second-order property of speaker's (non-neurophysiological?) mind. It is a "distinct level of things in the world", not to be explicated in causal terms. That an I-language is a second-order property of native speakers is not sufficient to secure the claim that facts about such competence states are uniquely linguistic. A second-order property is simply a property of the first-order properties of objects. For example, dispositional properties like solubility are second-order properties of objects. A salt crystal has the property of being soluble, and being soluble is a property of the first-order properties of the salt crystal. More specifically, being soluble is a property of the relative electro-static charge on sodium and chloride ions (a first-order property) in the presence of water molecules. But clearly, there is nothing uniquely linguistic about the second-order property of solubility.

Claims about the psychological reality of grammars, for Chomsky, are formulated as claims about which grammar a speaker uses. In a characteristic passage he writes:

Statements about the I-language are true or false, much the same way statements about the chemical structure of benzene . . . are true or false. The I-language L may be used by a speaker but not the I-language L', even if the two generate the same class of expressions. (1986:37)

The point that a grammar G is psychologically real in the sense that it is used by a speaker while the weakly equivalent G' is not is what makes the speaker of a language a rule-follower and rule-user and is what makes one description of a competence state psychologically real and the other not.

We shall explore how accepting Response C raises difficulties for the psycholinguist by considering the case of the Derivational Theory of Complexity [DTC] from the history of psycholinguistics. DTC was an early, if not the first, attempt to provide a performance model of the grammar outlined in ASPECTS OF THE THEORY OF SYNTAX with sufficient specificity and detail that Chomskian conjectures about the psychological reality of the so-called 'Standard Theory' could be tested (though some actual psycholinguistic uses of DTC preceded the publication of ASPECTS). DTC assumed the grammar (description of the I-language) of ASPECTS, and took on the assumption that the relation between that grammar and the parsing algorithm was transparent in the sense that the relation was isomorphic. The deep structure and surface structure of input strings was recovered by the parser which echoed the grammar, and the deep structure was derived from the surface structure by the application of 'inverse transformations'. DTC also assumed that each grammatical operation cost one unit of time and since the parser/grammar relation was one-to-one the temporal cost of constructing the deep structures from surface structures was the sum of the number of the applications of rules necessary for the derivation of the sentence. There are more subtle versions of DTC, but all versions share the notion of a relatively transparent relation between linguistic rules and

psychological operations. The motto was one temporal unit for each rule application.

Given the assumptions of DTC, the theory predicts that the mapping from deep to surface structure for active sentences requires one fewer rule applications than passives, so it would take less time to parse actives than passives. Some very early evidence appeared to confirm DTC (and the grammar of Aspects). However, later experiments "found no correlation between sentence processing time and the length of transformational derivation".⁸ There are three possible retrenched versions of DTC: reject the grammar, reject the assumption that the relation between the grammar and parser is isomorphic, or reject the computational complexity measure. Each alternative has subsequently been attempted, but it is the second avenue of retrenchment that shall be of particular interest to us here.

Fodor, Bever and Garrett (1974) were the first to suggest that the transparent, isomorphic relation between the grammar and parser be revised. In place of the isomorphic relation they substituted "heuristic strategies", with the effect that the subsequent performance model reduced the online computation involved in sentence comprehension. Now at one level, the rejection by Fodor et. al. of the transparency assumption is a standard piece of ordinary science. In the face of disconfirming evidence for the favored theory, hang onto the theory and reject an auxiliary hypothesis. However, the retrenchment in this case is self-defeating. If one permits the adoption of any "heuristic strategy" as the posited relation between the grammar and parser, then virtually any parser will model any set of rules. In this case, the parser/grammar relationship is completely unconstrained by the theory, and any sense in which the performance model can be used to test the psychological reality of the theory disappears.

Cognizant of the dangers heuristic grammar/parser relations pose for claims about the psychological reality of grammars, Berwick and Weinberg (1986) define a notion of grammatical covering in an attempt to respond to the problem. Informally characterized, they claim that one grammar G covers another G' if

- (1) both generate the same language $L(G) = L(G')$, that is the grammars are weakly equivalent; and (2) we can find parses of structural descriptions that G' assigns to sentences using G and then applying a "simple" or easily computed mapping to the resulting output. (1986:79-80)

⁸ Quoted in Berwick and Weinberg (1986:42) from Joan Bresnan, "A Realistic Transformational Grammar" in M. Halle, J. Bresnan and G. Miller (Eds.) *LINGUISTIC THEORY AND PSYCHOLOGICAL REALITY* (Cambridge: MIT Press, 1978).

They cash out "simple" as follows:

. . . [t]he usual definition of "simple" drawn from the formal literature is that of a string homomorphism. That is, if the parse of a sentence with respect to a grammar G is a string of numbers corresponding to the rules that were applied to generate the sentence under some arbitrary numbering of the rules of the grammar and some canonical mapping derivation sequence, the translation mapping that carries this string of numbers to a new string of numbers corresponding to the parse [under G'] must be a homomorphism under the concatenation. (1986:79-80)

If G covers G' , the grammars (or grammar and parser) are not merely weakly equivalent. The structure of the parse, but not the number of rule applications, is preserved under string homomorphism. The covering relationship, however, is weaker than strong equivalence and the transparency relation. In consequence, if computational cost measures are held constant, predictions of total time cost will vary radically from parser to parser.

Recall that Chomsky defines psychological reality in terms of the use of a grammar -- applying internally represented rules. However, if we follow Berwick and Weinberg's elegant suggestion, we accept a notion of grammatical covering on which the description of the I-language, e.g., G , is not strongly equivalent to the parser, e.g., G' , -- the used grammar. Once we do this, it is not easy to see what the idea of using the covering grammar comes to. By hypothesis, it is not the covering grammar, but G' , the parser, that is used. Worse yet, there are indefinitely many parsers that are covered by the competence grammar. Are we to suppose that the speaker can use only one of these, or many? If we suppose speakers may implement more than one of the covered parsers, then temporal costs will vary for the same grammar (holding temporal cost measures constant) and strikingly different predictions follow from the same competence theory. [What empirical content would attach to the claim that the speaker uses one, rather than many of the covered parsers?] In this case the grammar no longer provides a standard of performance in terms of temporal cost and loses its ability to function as one needs a competence theory to function in the performance model. On the other hand, if we take the speaker to use just one of the parsers covered by the grammar, then there is no advantage in claiming that the competence grammar is used, although the parse is structurally homomorphic to the parse of the competence grammar had it been used. A further difficulty is that, assuming that the covering relation is symmetric, any given G' (the parser) might cover indefinitely many competence grammars, G .

Berwick and Weinberg claim that loosening the relation between (competence) grammar and parser (performance grammar) provides the computational linguist with a performance model that has clear methodological

advantages. In response to the question "Why build a parser that is covered by a competence grammar?" they write:

The answer is that by keeping the levels of grammar and algorithmic realization distinct, it is easier to determine just what is contributing to discrepancies between theory and surface facts. For instance, if levels are kept distinct, then one is able to hold the grammar constant and vary the machine architectures to explore the possibility of good fit between psycholinguistic evidence and model. Suppose these results came to naught. We can then try to vary machine architecture and covering mappings, still seeking model and data compatibility. . . . [M]odularity of explanation permits modularity of scientific investigation. (1986:80)

As a methodological claim, the thesis is unassailable. If I understand Berwick and Weinberg correctly, their idea is that a modular conception of performance models facilitates the manipulation of elements of computational simulations of those models and promotes ease of identifying various models of the competence grammar. However, such methodological boons would issue from taking the covering grammar to identify a class of parsers. Berwick and Weinberg's response does not bear on the issue of what one could now mean by the psychological reality of the covering grammar. That distinguishing between a competence grammar (class of parsers) and a used parser is useful does not explicate what it could now mean to use the competence grammar. The use of a grammar, by hypothesis involves specific and determinate temporal costs, not a class of temporal costs for the same (grammatical) phenomenon.

Let us rehearse where we have been. The initial idea was that internally representing rules was a *sui generis* fact about Jones (Response C) that could be embraced to support the falsity of (2) in the skeptical paradox. In order to develop a performance model of a competence theory which is taken to describe those brute linguistic rule-facts, one must posit a grammar parser relation. If the relation is supposed to be transparent and consistent with DTC, we have a clear (but non-neutral) description of what it is to use those rules -- the transparency relation. But the transparency relation has empirical difficulties, so a string homomorphism was postulated to gain its methodological advantages. But now conceptual problems arise and philosophical difficulties resurface concerning what it means to say that a speaker uses the competence grammar (not the parsing grammar) and that the competence grammar is psychologically real. If we say that a class of parsers are the facts in virtue of which the speaker is in S, then there is incompatible psycholinguistic evidence about those facts that will *prima facie* both support that a speaker is and is not in S. For each parser will provide a different standard with respect to time costs for the same phenomenon. Alternatively, if we say to use G is to use just one G' covered by G, then it seems that we would be just as well off to say that G' is the real fact of the matter, and G a useful means of identifying G', but is neither used by the speaker nor psychologically real. This latter

suggestion has been made by a number of theorists and rejected by Chomsky (see, for example, Soames 1984).

I have argued that if the linguist attempts to avoid the skeptical paradox that Kripke attributes to Wittgenstein along the lines Chomsky suggests in *KNOWLEDGE OF LANGUAGE*, then either he ends up rejecting the performance/competence distinction or (given the methodological requirements on a performance model) the paradox resurfaces for the psycholinguist in terms of saying what it is to use the competence grammar. If what I have argued is correct, then the alternative strategies open to the linguist for avoiding the paradox are to either claim Response D, that there is nothing about a speaker in virtue of which he uses one grammar rather than another (i.e., linguists' competence state attributions are not fact stating), or Response B, that facts about competence states do not justify grammar attributions. Neither of these alternatives is acceptable, for both require considerable revision in the typical conception of competence. At least this is the situation unless the linguist can show (or plausibly hope) that there is a description of the facts that constitute what it is to use a rule in neutral terms. I have not, of course, addressed the issue of whether a neutral description of rule-following can be articulated by computational theories of mind. That is a distinct and very long story.

While I do not think the skeptical paradox poses insuperable conceptual difficulties for generative linguistics (even if one does not embrace the language of thought hypothesis and the computational metaphor), I have attempted to show that there is a troublesome conceptual problem here. Surely one of the conceptual burdens of psycholinguistics is to say what constitutes the use of rules. The Wittgensteinian skeptical paradox makes explicit the conceptual difficulties inherent in that task.

Acknowledgements

This paper has been vastly improved by the comments and sage editorial advice of Wayne Cowart. I am deeply indebted to him. I would also like to thank David McCarty for helpful comments on a much earlier draft.

References

- Berwick, R. and A. Weinberg. (1986). *The Grammatical Basis of Linguistic Performance*. Cambridge, MA: MIT Press.
- Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.
- Fodor, J. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.
- _____. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

- Kripke, S. (1984). *Wittgenstein On Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Soames, S. (1984). "Linguistics and Psychology". *Linguistics and Philosophy* 7.2. MIT Press.