

**Knowledge of
Language:
Its Nature, Origin, and Use**

Noam Chomsky

CONVERGENCE

**A Series Founded, Planned, and Edited
by Ruth Nanda Anshen**

PRAEGER

Westport, Connecticut
London

as (288), where both the verb and the preposition play a role; see Marantz (1984). A number of other questions have also been left unsettled, even for the restricted case of English.

140. Recall that (171) was derived apart from (170), and that it holds only for CHAINS consisting of A-positions; see p. 178.

4

Questions about Rules

4.1. SOME SKEPTICAL DOUBTS

We have, so far, considered the first two of the questions (1) of Chapter 1 that express the essential research program in generative grammar: (1i) what constitutes knowledge of language, and (1ii) how does it arise? As for (1i), to know the language L (an I-language) is to be in a certain state S_L of the language faculty, one of the components of the mind/brain. L , the language known (or had, or internalized) in the state S_L , is a system with two components: core and periphery. There is a fixed initial state S_0 of the language faculty consisting of a system of principles associated with certain parameters of variation and a markedness system with several components of its own. The proposed answer to (1ii) is that the state S_L is attained by setting parameters of S_0 in one of the permissible ways, yielding the core, and adding a periphery of marked exceptions on the basis of specific experience, in accordance with the markedness principles of S_0 . The core, then, consists of the set of values selected for parameters of the core system of S_0 ; this is the essential part of what is "learned," if that is the correct term for this process of fixing knowledge of a particular language. The grammar of the language L is the linguist's theory of L , consisting of a core grammar and an account of the periphery.

These proposed answers to the questions (1i) and (1ii) are presented at a certain level of abstraction from (largely unknown) mechanisms. This level of abstraction is appropriate in that by moving to it, we are able to discover and formulate explanatory principles of some significance, principles that should, furthermore, help guide the search for mechanisms.

The result of this process of parameter determination and periphery formation is a full and richly articulated system of knowledge. Much of what is known lacks relevant grounding in experience, justification, or good reasons and is not derived by any general reliable procedures. The same may well be true of large areas of what might be called "commonsense knowledge and understanding," and if the model of Peircean abduction is correct, it may also be true, to a significant extent, of scientific knowledge as well, although in this case argument and evidence are required to justify knowledge claims.¹ Along these lines, we can develop a possible answer to our variant of "Plato's problem."

I have so far said little about question (1iii), how knowledge of language is put to use, apart from a few remarks about the perceptual aspect of this problem. With regard to the far more obscure production aspect, the commonsense answer is that use of language is rule-guided behavior: We have (generally tacit) knowledge of the rules of language and we use them in constructing "free expressions" in Jespersen's sense. Some questions have been raised as to whether the concept of rule in the conventional sense is an appropriate one for the elements of language. Let us put these aside for the moment and assume that it is indeed legitimate to regard the language as a rule system or to project a rule system in some manner from the language that a person knows. Let us assume, then, that the commonsense picture is more or less correct, with the modifications already discussed: in particular, the abstraction from the sociopolitical and normative-teleological elements of the commonsense notions of language and rule (see Chapter 2).

An attempt to provide some substance to this commonsense account runs into numerous problems, some of them classic ones. In the first place, there are what we might call the "Cartesian problems." In the Cartesian view, the "beast-machine" is "compelled" to act in a certain way when its parts are arranged in a particular manner, but a creature with a mind is only "incited or inclined" to do so because "the Soul, despite the disposition of the body, can prevent these movements when it has the ability to reflect on its actions and when the body is able to obey" (La Forge). Human action, including the use of rules of language, is free and indeterminate. Descartes believed that

these matters may surpass human understanding: We may not "have intelligence enough" to gain any real understanding of them, although "we are so conscious of the liberty and indifference which exists in us that there is nothing that we comprehend more clearly and perfectly" and "it would be absurd to doubt that of which we inwardly experience and perceive as existing within ourselves just because we do not comprehend a matter which from its nature we know to be incomprehensible." One can question various aspects of this formulation: for example, that we literally "know" the matter to be incomprehensible, and that the limits are not merely those of human intelligence but rather of undifferentiated mind, not part of the biological world at all. But even so, it is difficult to avoid the conclusion that serious problems are touched on here, perhaps impenetrable mysteries for the human mind, which is; after all, a specific biological system and not a "universal instrument which can serve for all contingencies," as Descartes held in another context. There is no more reason to suppose humans to be capable of solving every problem they can formulate than to expect rats to be able to solve any maze.²

A second class of problems concerning rule following are what we might call the "Wittgensteinian problems."³ This topic has been greatly clarified by Saul Kripke's recent exegesis and analysis (Kripke, 1982). I will not enter into the textual question of whether Kripke's version of Wittgenstein is the correct one, but will merely assume that it is and will refer to Kripke's Wittgenstein henceforth as "Wittgenstein"; the quotes below are from Kripke, unless otherwise indicated. Kripke does not specifically endorse the picture he presents, but it is undoubtedly an influential and important one, and one that appears to be highly relevant to the concerns of generative grammar, as Kripke stresses several times. Of the various general critiques that have been presented over the years concerning the program and conceptual framework of generative grammar, this seems to me the most interesting.

Kripke suggests that "our understanding of the notion of 'competence' [equivalently, 'knowledge of language,' as used above] is dependent on our understanding of the idea of 'following a rule,'" so that Wittgenstein's skeptical paradox concerning rule following crucially bears on the central questions

addressed in generative grammar. Furthermore, if we accept Wittgenstein's solution to his skeptical paradox, then

the notion of 'competence' will be seen in a light radically different from the way it implicitly is seen in much of the literature of linguistics. For *if* statements attributing rule following are neither to be regarded as stating facts, nor to be thought of as *explaining* our behavior... [as Wittgenstein concludes], it would seem that the *use* of the ideas of rules and of competence in linguistics needs serious reconsideration, even if these notions are not rendered 'meaningless.'

One aspect of the account given earlier, and of the work reviewed, is that it is presented in the framework of individual psychology: Knowledge of language (competence) is taken to be a state of the individual mind/brain. Wittgenstein's solution to the skeptical paradox concerning rule following is crucially framed in terms of a community of language users. Furthermore, the preceding account assumed that the statements of grammar and UG are not different in principle from the statements of natural science theories; they are factual, in whatever sense statements about valence or chemical structure or visual processing mechanisms are factual and involve truth claims. We can look forward to the day when these statements will be incorporated into a broader theory concerning mechanisms which will explain why they are true (or why they are not) at the level of abstraction at which they are formulated. But all of this appears to be undermined by Wittgenstein's solution to his skeptical paradox. As Kripke puts it, generative grammar "seems to give an explanation of the type Wittgenstein would not permit." Thus, "Depending on one's standpoint, one might view the tension revealed here between modern linguistics and Wittgenstein's sceptical critique as casting doubt on the linguistics, or on Wittgenstein's sceptical critique—or both." He further observes that the issue has nothing to do with whether rules are explicitly stated—with whether people have access to the rules that constitute their knowledge, in our terms. Note also that the questions arise even under the abstraction from the sociopolitical and normative-teleological aspects of the commonsense notion of language, that is, under the idealizations we have assumed.⁴

Wittgenstein's skeptical paradox, in brief, is this. Given a rule R, there is no fact about my past experience (including my conscious mental states) that justifies my belief that the next application of R does or does not conform to my intentions. There is, Wittgenstein argues, no fact about me that tells me whether I am following R or R', which coincides with R in past cases but not future ones. Specifically, there is no way for me to know whether I am following the rule of addition or another rule (involving "quus," not "plus") which gives the answer 5 for all pairs beyond the numbers for which I have previously given sums; "there was no *fact* about me that constituted my having meant plus rather than quus," and more generally, "there can be no such thing as meaning anything by any word." Each application of a rule is "a leap in the dark." My application of a rule "is an unjustified stab in the dark. I apply the rule *blindly*." The argument is not limited to use of concepts but extends to any kind of rule application.

In short, if I follow R, I do so without reasons. I am just so constituted. So far, these conclusions offer no serious challenge to the account discussed earlier. I follow R because S_0 maps the data presented into S_L , which incorporates R; then "I apply the rule R blindly." There is no answer to Wittgenstein's skeptic and there need be none. My knowledge, in this instance, is ungrounded. I know that $27 + 5 = 32$, that this thing is a desk, that in a certain sentence a pronoun cannot be referentially dependent on a certain noun phrase, and so forth, as a consequence of knowing rules, which I follow (or I may not for some reason, perhaps by choice, thus giving wrong answers). But I have no grounds for my knowledge in any useful general sense of the term and no reasons for following the rules: I just do it. If I had been differently constituted, with a different structure of mind/brain (S'_0 instead of S_0), I would have come to know and follow different rules (or none) on the basis of the same experience, or I might have constructed different experience from the same physical events in my environment.

The apparent problem for our account arises when we consider a different question: How can I tell whether you are following R or R'? Under what circumstances does it make sense for me to attribute rule following to you? When is this attribution correct or justified? Here we may distinguish two

cases: my doing so as a person in ordinary life, and my doing so as a scientist seeking to discover the truth about the language faculty. The first case raises a question of description: When do I, in fact, attribute to you a particular instance of rule following? Both cases raise questions of justification: When am I entitled, as a person in ordinary life or as a scientist, to say that you are following a rule?

Consider the first case: ascription of rule following in ordinary life. Wittgenstein holds that I am entitled to say that you are following R if you give the responses I am inclined to give and you interact properly with my community, and if the practice of attributing the rule R to you has a role and utility in our communal life. Then I "take you into the community" to which I belong. The community attributes a concept (rule) to an individual as long as he or she conforms to the behavior of the community, its "form of life." Deviant behavior is rare as a matter of "brute fact"; hence, this practice of attributing concepts and rules is a useful one. Because attribution of rule following requires reference to the practices of a community, there can be no "private language." There is no substance or sense to the idea of a person following a rule privately. It seems that the "individual psychology" framework of generative grammar is undermined.

Wittgenstein holds, then, that "If we confine ourselves to looking at one person alone, his psychological states and his external behavior, this is as far as we can go. We can say that he acts confidently at each application of a rule... there can be no facts about him in virtue of which he accords with his intentions or not." "If one person is considered in isolation, the notion of a rule as guiding the person who adopts it can have *no* substantive content," so that the statements of a generative grammar, which appear to consider a person in isolation, can have no substantive content. But "The situation is very different if we widen our gaze from consideration of the rule follower alone and allow ourselves to consider him as interacting with a wider community. Others will then have justification conditions for attributing correct or incorrect rule following to the subject..." namely, if his responses coincide with theirs. There are no truth conditions for "Jones is following rule R," because there is no fact of the matter; and more generally, we should not seek truth

conditions for the expressions of ordinary language but rather assertability conditions.

As for the assertability conditions, Jones is entitled to say "I mean addition by 'plus,'" subject to various provisos, "whenever he has the feeling of confidence... that he can give 'correct' responses in new cases." His inclination to go on in a certain way is to be regarded as "primitive." Smith is entitled to say that Jones means addition by "plus" if he judges that Jones is inclined to give the answers to addition problems that he, Smith, is inclined to give; and since as a matter of brute fact the community is roughly uniform in its practices, this "game" of attributing rule following has a role and utility in our lives. Smith's behavior too is "a primitive part of the language game."

Recall that Wittgenstein's solution is not intended to reform language use, but to describe it, to show why it is fine as it is. It must therefore be descriptively adequate. But this account is very far from descriptively adequate; it simply does not work for standard cases of attribution of rule following. Possibly, the discussion is obscured by concentrating on cases that are felt to be deep in their character and implications, and that certainly are deeply embedded in the philosophical tradition, specifically, attribution of concepts. These are, furthermore, cases where there is understood to be some normative standard of correctness. Let us consider, however, typical cases of attribution of rule following that are less "loaded" in this sense.

At a certain stage of language growth, children characteristically overgeneralize: They say *sleeped* instead of *slept*, *brang* (on the analogy of *sang*) instead of *brought*, and so forth. We have no difficulty in attributing to them rules for formation of past tense, rules that we recognize to be different from our own. In this case, 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 commonsense notion of language. If all adults were to die from some sudden disease, the "language would change" and these irregularities would be erased. The child's rule would now be "correct" for the new language. In accordance with the move suggested earlier, we may, then, say that the child is following a rule of his or her language at the time, one of the possible human languages, but not exactly ours.

To avoid the issue of the normative-teleological aspect of the commonsense notion, consider a different case. Suppose that we have visitors from a dialect area different from ours where lax and tense /i/ are merged before /g/, so that the words *regal* and *wriggle* are pronounced the same way, with an intermediate vowel; or where people say "I want for to do it myself" or "he went to symphony" instead of "I want to do it myself" and "he went to the symphony." Again, we would say that they are following rules, even though their responses are not those we are inclined to give, and we do not take them into our linguistic community in these respects. They do not share our "form of life" or interact with our community in the relevant sense. In such cases there is no question of "correctness" any more than in the choice between English and French. Furthermore, our conclusion that they are following rules different from our own has no obvious role or utility in our lives, although the conclusion is more likely to be drawn in this case than in the case of conformity to our practice; under the latter circumstances, such questions are commonly ignored. It might be that the usual case of attribution of rule following is when the responses do *not* accord with ours, when they are unexpected and unfamiliar. Few people other than linguists would be inclined to say that Jones is following condition (B) of the binding theory when he understands *them* to be free in reference, not dependent on *the men*, in "the men expected to like them." This case, although unusual in practice, does follow the Wittgensteinian paradigm; normal cases do not.

The same applies to attribution of concepts. Like many people, I learned the word *livid* from the phrase *livid with rage*. In my language at the time, it meant something like "flushed" or "red." Later, my linguistic knowledge and practice changed, and in my current language it means something like "pale." I have no difficulty in attributing a different rule (my earlier one) to someone who I see follows it. Similarly, it is standard to attribute concepts different from ours to children and foreigners, or speakers of other languages. In the *plus-qui-us* case, we would, as players of the normal language game, attribute one or the other concept to people by inspecting their behavior, although in one case their responses would not accord with our own. There may be a question as to *how* we do it, but there seems

little doubt that we do do it. Furthermore, none of this seems to have much if any utility in our lives.

In standard cases of attribution of rule following, such as those mentioned, the rules in question may or may not be followed in behavior. The child who overgeneralizes, for example, may choose not to apply his or her rule for forming the past tense of *sleep* in some particular case, or might not follow the rule for some other reason (and might, perversely, even say *slept*, violating the rule). Our visitors might pronounce *regal* and *wriggle* with a tense-lax vowel distinction (as we do), possibly by choice, thus violating what is their rule at the time (but keeping this rule, though violating it), and so forth. Thus, even when we drop any of the various normative considerations, the rules are not descriptions of behavior or of regularities in behavior (in principle, our visitors might choose to violate their rule most or all of the time, for one reason or another). The problem of determining when the rule is being followed, and when not, may be a difficult empirical one, but there seems little doubt that it does arise in the manner just indicated.

One of the centerpieces of Kripke's discussion is paragraph 202 of the *Philosophical Investigations*:⁵

... to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately'; otherwise thinking one was obeying a rule would be the same thing as obeying it.

This passage misconstrues our attribution of rule following in ordinary language or science, and the argument it presents is not valid. The premise is correct: Jones may or may not be following a rule whatever he is thinking, either because he doesn't think about rules at all or because his self-analysis is wrong for one reason or another (in general, the account that people give for their behavior is highly unreliable, even when they feel that they can offer one). It is correct, then, that thinking one is obeying the rule is not the same thing as obeying it. But the conclusion does not follow. If we say that Jones is obeying a rule "privately," and hence that he is obeying a rule, nothing follows at all about whether he thinks that he is obeying the rule. In fact, we do say that Jones obeys a rule "privately"—that

is just the way we play the game and no reason has been advanced as to why we should do otherwise—even if he thinks he is obeying a different rule or has no idea about rule following (and is responding differently than we would). Correspondingly, we must avoid the temptation to assume some notion of “accessibility to consciousness” with regard to mental states and their contents. This seems inconsistent with the way we use the relevant concepts in normal discourse, however deeply rooted the assumption may be in various philosophical theories, and it is an assumption that can only stand in the way of a descriptively adequate or genuinely explanatory theory of mind.

At the very end of his discussion (p. 110), Kripke brings up a case that might be construed as being of the kind discussed above, where attribution of rule following violates the Wittgensteinian paradigm: the case of Robinson Crusoe, not part of any community. Kripke asks whether the Wittgensteinian argument against the possibility of a “private language” entails “that Robinson Crusoe, isolated on an island, cannot be said to follow any rules, no matter what he does,” referring to a passage where Wittgenstein discusses the “somewhat similar question” of a person playing a familiar game translated by some rule into a different modality. If Robinson Crusoe’s responses are those we would be inclined to give, then this case raises no new questions; it is essentially the same as the case of our meeting people whose responses agree with our own, so that we attribute rule following to them in accordance with the Wittgensteinian paradigm, which as Kripke outlines it does not ask whether they are part of a community but rather whether we can take them into our community. The case becomes interesting, however, if Robinson Crusoe gives responses different from ours, that is, speaks a language of his own, shared by no community, in particular not our community. If the case is understood this way, then it does serve as a—rather exotic—example of the type discussed earlier. Let us interpret Kripke’s discussion to include this case—noting, however, that this may not be what was intended—and ask how his account, so construed, applies to the cases discussed earlier (quite normal cases, so it seems).

Kripke argues that we can still attribute rule following to Robinson Crusoe in accordance with the Wittgensteinian solution. Namely, we regard him as a *person* who acquires rules

under certain experiences, although not our rules, because we had different experiences. Then we can take him into the broader community of persons, who share our “form of life” in a broader sense. “Our community can assert of any individual that he follows a rule if he passes the tests for rule following applied to any member of the community,” that is, if he acts in the manner of rule follower although he doesn’t give our responses. This would include the cases discussed earlier but at the cost of abandoning any consequences of the “private language argument” that bear on the attribution of rules within the framework of individual psychology.

Consider more closely the statement that “Our community can assert of any individual that he follows a rule if he passes the tests for rule following applied to any member of the community.” Assuming that Robinson Crusoe passes the tests for rule following in the community of persons, we can say that “he follows a rule.” But which rule does he follow? Here, the Wittgensteinian paradigm is no help. The extension of the Wittgensteinian paradigm to the Robinson Crusoe case is inadequate to the task at hand, even if we accept it as legitimate.

Recall Wittgenstein’s solution to his skeptical paradox: “The situation is very different if we widen our gaze from consideration of the rule follower alone and allow ourselves to consider him as interacting with a wider community. Others will then have justification conditions for attributing correct or incorrect rule following to the subject...,” namely, if his responses coincide with theirs. But Robinson Crusoe does not interact with the wider community of persons to which we assign him on the basis of his behavior. Therefore, the Wittgensteinian solution does not apply to the Robinson Crusoe case: As formulated, it does not permit us to consider Robinson Crusoe to be a rule follower, because he does not interact with the community of persons, and it clearly does not permit us to determine which rule he is following. The first defect might be overcome by modifying the Wittgensteinian paradigm along the lines of the interpretation of Kripke’s remarks sketched above, but the second is a defect of principle. Recall that this is a perfectly normal case in which we attribute not only the general property of rule following but also the following of particular rules. The defects in the analysis are therefore rather serious.

Furthermore, there seems to be a crucial equivocation in the concept "form of life," which plays a central role in the argument just sketched. The term is defined (by Kripke) to refer to "the set of responses in which we agree, and the way they interweave with our activities" (p. 96). In this sense, I take you into the community sharing my "form of life" if your responses are like mine, in accordance with the Wittgensteinian paradigm for attribution of rule following. But in this sense, Robinson Crusoe, in the case we are considering, does not share our "form of life," and the solution collapses if intended to capture normal usage. We cannot attribute rule following to Robinson Crusoe, or in standard cases of the sort mentioned earlier. But Kripke also suggests a metaphorical usage of the phrase "form of life." In this extended sense, the "form of life" (he gives the term in quotes, indicating that it is a metaphorical extension) refers to the "highly species-specific constraints" that "lead a child to project, on the basis of exposure to a limited corpus of sentences, a variety of new sentences for new situations" (p. 97n). Here, the "form of life" refers to characteristic species behavior. It is this sense that is relevant to attribution of rule following or possession of concepts when the behavior does not match our own. Robinson Crusoe shares our "form of life" in this extended sense, although we need some further method to determine which rules he follows, which concepts he uses.

In the terms of the earlier discussion, the distinction is one of level of description: The technical usage of "form of life" is at the level of particular grammar (the attained language); in the extended sense it is at the level of UG (S_0). We might modify the Wittgensteinian solution to incorporate this distinction explicitly, so that it begins to approach normal usage. If we do, however, we derive a very different analysis of the "practice" of attributing concepts and rule-governed behavior, one that undermines the private language argument and the consequences drawn from it. A member of the species might well have unique experience that yields a unique rule system, a private language, although we could "take him into our community" in the broader sense of "form of life."

Indeed, this is not only the normal case but arguably the only case, if we investigate a person's language in sufficient detail. That is, we can expect Jones's language to be different from ours in at least some respects, and a correct analysis of rule

following should be holistic, accounting for Jones's following of a particular rule not in isolation but against the background of (tacit or explicit assumptions about) his whole language and no doubt much more. This conclusion, in fact, is implicit in the Wittgensteinian approach. In these broader respects we can hardly expect Jones to be like us.

Returning to the statement that "If one person is considered in isolation, the notion of a rule as guiding the person who adopts it can have *no* substantive content" (p. 89)—the conclusion that seemed to undermine the individual psychology framework of generative grammar—we see that this must be understood as referring not to an individual whose behavior is unique but to someone "considered in isolation" in the sense that he is not considered as a person, like us. But now the argument against private language is defanged. We consider Robinson Crusoe to be a person, like us. He has a private language with its own rules, which we discover and attribute to him by some means other than those allowed in Wittgenstein's solution to the skeptical paradox.

Note that we might also say that if a sample of water is "considered in isolation," not regarded as water, then we can say nothing about its chemical constitution and so forth; and we can say nothing about a fruit fly in a genetics experiment if we do not consider it as a member of the class of fruit flies. An entity becomes an object of inquiry (scientific or commonsense) only under a particular description, only insofar as it is assigned to a particular natural kind. We investigate a particular thing (already tacitly assuming some framework of description and understanding in accord with which it is a thing), tentatively decide that it is a sample of water, and then learn about its properties by investigating it and other samples. Similarly, we decide that Robinson Crusoe is a person on the basis of certain of his characteristics, and then determine the rules of his language by investigating his behavior and reactions and those of others, with different rules, who belong to the same natural kind. But there are no interesting conclusions to be drawn from this, beyond those that hold of descriptive commentary (scientific or otherwise) quite generally and hence are not relevant here.

We may ask how, in ordinary life, we assign Robinson Crusoe to the category of persons and what the sense of this attribution is; and whether, as scientists, we are entitled to say

that this attribution amounts to a factual claim that Robinson Crusoe shares with other persons some actual property—specifically, the initial state S_0 of the language faculty—so that given his experience he follows the rules of the attained state S_L , not our rules. The answer to the first question seems to be an intuitive and vague version of the scientist's answer: To be a person is to be an entity of a certain kind, with certain properties; we decide that Robinson Crusoe is an entity of this type, with these properties of personhood, by investigating what he does under various conditions. The status of these judgments becomes clearer when we consider the second question, to which we turn directly.

Keeping to the first question, we are led back, I think, to something like the Cartesian conception of "other minds." According to this view, I attribute to you a mind like mine if you pass tests that indicate that your behavior exhibits intelligence,⁶ where to "exhibit intelligence" is to exceed the limits of mechanism in the sense of Cartesian contact mechanics. If a variety of such experiments succeed in showing this, then "I would not be reasonable if I did not conclude that [the subjects] are like me," possessing a mind (Cordemoy). The major tests outlined by the Cartesians, including Descartes himself, involve what I have called elsewhere "the creative aspect of language use," the use of language that is unbounded, stimulus-free, appropriate to situations, coherent, and evoking appropriate thoughts in me. Or if your behavior reflects understanding and the exercise of will and choice as distinct from mechanical response, then I attribute to you possession of mind, a power that exceeds the bounds of Cartesian contact mechanics (as does, in fact, the motion of the heavenly bodies, as Newton subsequently showed, thus undermining the mind-body problem as Descartes formulated it since there is no longer any clear content to the notion of body). By various tests, I try to determine whether your "cognoscitive power" is "properly called mind," that is, whether it is not "purely passive" but rather "forms new ideas in the fancy or attends to those already formed," not completely under the control of sense or imagination or memory (*Rules for the Direction of the Mind*). If so, I "take you into my community" in the broader sense; I take you to be a person, sharing my "form of life" in the metaphorical sense (at the level

of UG), and I assume that you follow rules as I would under similar past/present conditions, although the rules I attribute to you are not mine—our responses differ, and we do not share a "form of life" in the technical sense. There is no issue of "utility" in all of this. I just make these determinations, without reasons, just as I follow rules, without reasons, as a reflection of my nature.

The Kripke-Wittgenstein dismissal of the Cartesian position is based on Humean arguments (namely, that we have no impression of self) that do not seem to me to bear on the relevant aspects of Cartesian thought, because they fail to take into account the respects in which attribution of mind is a theoretical move, part of a pattern of explanation based on assumptions with regard to the limitations of mechanics. We surely need not accept Cartesian mechanics or metaphysics or the beast-machine conclusions, nor need we accept the model of conscious testing and theory construction for normal usage. But the kernel of thinking seems plausible enough, and it is not unlike the extension of the Wittgensteinian paradigm, considered above, to the cases that clearly violate the formulation in terms of "interaction with a wider community" and the technical notion of "form of life." On the Cartesian assumptions, I attribute to you rules (though perhaps not mine), rules that I would have followed had I had your experience. I do this because you seem to be a person, exhibiting characteristic features of will and choice, the creative aspect of language use, and other indications of intelligence (and for Cordemoy, looking like me). Further analysis, however, is required to explain how I assign particular rules to you.

Summarizing, Kripke's Wittgenstein holds that

- (I) "To judge whether an individual is indeed following a given rule in particular applications" is to determine "whether his responses agree with their own."
- (II) We therefore reject the "'private model' of rule following," according to which "the notion of a person following a given rule is to be analyzed simply in terms of facts about the rule follower and the rule follower alone, without reference to his membership in a wider community."

- (III) "Our community can assert of any individual that he follows a rule if he passes the tests for rule following applied to any member of the community."

As for (I), it is not true in standard cases. We regularly judge that people are following rules when their responses differ from our own. As for (III), it is tenable if we understand it to mean that whether or not an individual's "responses agree with [our] own," we may assert that he or she follows rules if he or she passes the tests for rule following, not with respect to particular rules or with reference to any particular community of rule users, but more generally: He or she acts as a person, passing the tests for "other minds" in roughly the Cartesian sense (with the provisos noted). By virtue of such facts about the individual (which are not facts about the individual's experience and mental life), we "take him into the community" of persons and assume him to be following rules like ours, though perhaps not our rules. It remains to determine what these rules are by observation, applying our own intuitive methods and criteria, whatever they may be—evidently not those of the Wittgensteinian paradigm. All of this is done without reasons, just as we follow rules ourselves without having reasons ("blindly"). Contrary to (II), there seems nothing objectionable about the "private model" of rule following, nor is any serious alternative proposed to it, at least in any sense relevant to the explanations and concepts involving "competence" or "knowledge of language" in generative grammar; reference to a community of users of a language seems beside the point.

All of this has to do with the way, as persons, we attribute rule following without much if any reflection to others we take to be persons (and also, probably, to nonpersons in some cases). But it does not yet deal with the objection that there is no fact of the matter. This issue arises when we consider the second question raised earlier: What about our conclusion, as scientists, that Jones is following the rule R? Here we need reasons and justification. Can they be given?

The approach sketched earlier holds that we should proceed as follows. We amass evidence about Jones, his behavior, his judgments, his history, his physiology, or whatever else may bear on the matter. We also consider comparable evidence

about others, which is relevant on the plausible empirical assumption that their genetic endowment is in relevant respects the same as his, just as we regard a particular sample of water as water, and a particular fruit fly as a fruit fly. We then try (in principle) to construct a complete theory, the best one we can, of relevant aspects of how Jones is constructed—of the kind of "machine" he is, if one likes.

One heavy empirical condition is that this theory must incorporate a theory of the initial state that suffices to yield both the account of Jones's language (given relevant experience) and the account of the state attained by others (given their different experience). This theory is about Jones's capacities and how they are realized, these being facts about Jones. At the same time it is a theory about persons, the category to which we take Jones to belong as an empirical assumption.

Suppose that our best theory takes the initial state to incorporate as one component the initial state S_0 of the language faculty (a distinct component of the mind/brain), certain processing mechanisms, a certain organization and size of memory, a theory of random errors and malfunctions (parts wearing out or whatever), and so forth, all of this as a species characteristic. This theory provides an account of the current state of the person as incorporating a particular language L, which is a particular realization of the principles of S_0 with values of parameters fixed (its core) and a periphery added. Then we conclude that the person follows the rules of L or those projected from it,⁷ which determine what expressions mean for him, their correct forms for him, and so forth. This approach is not immune to general skeptical arguments—inductive uncertainty, Hilary Putnam's antirealist arguments, and others. But these are not relevant here, because they bear on science more generally. It is not clear that there are any further skeptical arguments that apply. A particular theory of this sort may certainly be wrong and may be shown to be wrong, for example, if the theory UG of S_0 proposed to account for Jones's language (and, thus, to explain facts about Jones's judgments and behavior, in the explanatory model discussed earlier) fails with respect to someone else, say a speaker of Japanese. In fact, this has repeatedly been the case and is surely the case with regard to current theories. It seems clear, then, that these theories are empirical ones, which could also be right.

Kripke argues against a "dispositional" account of rule following and concludes that the account must be "normative," not "descriptive" (p. 37). As he notes, the preceding account is not dispositional (it says little about what a person is disposed to say under particular circumstances), and it is also not "causal (neurophysiological)." Furthermore, the account is not "functionalist"; it does not "regard psychology as given by a set of causal connections, analogous to the *causal* operations of a machine,"⁸ although it has causal aspects: namely, with regard to the apparently deterministic passage from S_0 to the attained state S_L , and (at least in part) with regard to the operations of a parsing mechanism that uses the language. But the account of "competence" is descriptive: It deals with the configuration and structure of the mind/brain and takes one element of it, the component L , to be an instantiation of a certain general system that is one part of the human biological endowment. We could regard this instantiation as a particular program (machine), although guarding against implications that it determines behavior. Thus, an account can be descriptive although it is neither dispositional nor causal (neurophysiological or functional), in Kripke's sense.

Kripke argues, however, that what program a machine is following is not an objective fact about the machine, and that we can distinguish between a machine's malfunctioning and its following its program only in terms of the intention of the designer: "Whether a machine ever malfunctions and, if so, when, is not a property of the machine itself as a physical object but is well defined only in terms of its program, as stipulated by its designer." If a machine fell from the sky, there would be no answer to the question: "What program is it following?"

In our case there is no designer, but, nevertheless, we do assert that the machine incorporates a particular program. This assertion is part of a more general account of the properties of the mind/brain, an account that defines "malfunction" and "intrusion of extraneous factors," and is answerable to a wide range of empirical evidence, including evidence concerning the person's history and concerning speakers of other languages, and in principle much else: physiology, psychological experiment, brain damage, biochemistry, and so forth. Our assumption is that the person before us has a language with particular rules and principles along with other systems that interact with

it as a matter of mental/physiological fact, and which we might think of as a particular machine program with a particular data structure, and so forth. There may be empirical problems in sorting out the effects of these interacting systems, but these seem to be problems of natural science. In this world, with its regularities, the problems do not seem hopelessly difficult. Indeed, they have been addressed with some success.

It should furthermore be noted that there is no necessary restriction to human behavior here. It might be appropriate to describe the way a sheep dog collects the flock, or the way a spider spins a web, or the way a cockroach walks in terms of rule following, with reference to an underlying "competence" consisting of a system of rules of some sort—a faculty that might be intact although unusable, or might be misused for some reason in particular cases, and that underlies abilities that might be impaired, lost, recovered, or whatever.

Although the matter is not strictly relevant here, it seems to me that Kripke's conclusions about machines of the usual sort are too strong. Suppose that a machine fell from the sky, say an IBM PC with a particular operating system and a particular program stored in machine memory. Could we distinguish hardware, operating system, particular program? It seems that we could learn something about the matter by investigating input-output properties. For example, we might ask what aspects of the machine's functioning can be affected just through use of the keyboard and what can be changed by inserting a new board, or by going inside the microprocessor and manipulating the circuitry, and so forth. We could distinguish properties of this specific device from those that hold of any device made up of such components (properties of random behavior or distribution of output). We could develop a theory of the machine, distinguishing hardware, memory, operating system, program, and perhaps more. It is hard to see how this would be crucially different in the respects relevant here from a theory of other physical systems, say the interior of the sun, an internal combustion engine, or the organization of neurobehavioral units (reflexes, oscillators, and servomechanisms) that explain how a cockroach walks.⁹

Wittgenstein's skeptic goes beyond his argument when he concludes that there are no facts of the matter. What he has shown is that the facts concerning Jones's past behavior and

conscious mental states are not sufficient to establish that Jones is following the rule R, but it does not follow that "there can be no facts about him in virtue of which he accords with his intentions or not," that is, in virtue of which he can be said to follow the rule R. Kripke notes that there might be a neurophysiological theory that would explain a person's behavior, but this would not be to the point because it does not have the required prescriptive force: It does not provide justification and, thus, does not answer the skeptic; and furthermore, such theories would not be relevant to ascription of rule following by others who know nothing of these matters but do ascribe rule following. But it does not follow that we must accept the skeptical conclusion that there is no fact as to whether Jones means plus or quus, or whether he follows the rules of binding theory, or the rule that merges tense and lax /i/ before /g/. The approach just outlined leads to confirmable theories as to whether indeed Jones follows these rules.

The entire discussion is, in fact, a familiar one. The approach we have outlined is a variant of what Richard Popkin (1979) calls the "constructive scepticism" developed by Mersenne and Gassendi in response to the skeptical crisis of the seventeenth century, their "new outlook, ... doubting our abilities to find grounds for our knowledge" and recognizing that "the secrets of nature, of things-in-themselves, are forever hidden from us," while "accepting and increasing the knowledge itself"—a position based on "the recognition that absolutely certain grounds could not be given for our knowledge, and yet that we possess standards for evaluating the reliability and applicability of what we have found out about the world"; essentially, the standard outlook of modern science.

Wittgenstein holds that there is a crucial difference between the case of Robinson Crusoe (or any rule follower) and the case of a molecule of water or benzene. In the latter case, we regard a particular entity as a sample of water or benzene and then determine its properties, which are real properties of that sample. In the case of Robinson Crusoe, we regard him as a person and thus consider him a rule follower who behaves as we would had we had his experiences; and somehow, though not by the Wittgensteinian paradigm, we identify particular rules that he follows—commonly, not ours. In our terms, we assume that he

has a language faculty that shares with ours the state S_0 and attains a state S_L different from ours, on the basis of which we can develop an account for his current perceptions and actions. But, Wittgenstein holds, we are not entitled to go on to assert that the initial state S_0 or the attained state of knowledge S_L are real properties of this individual, that statements about them are true or false. There is no fact of the matter about Robinson Crusoe, or about persons in general, apart from the facts about communities. However, his arguments fall far short of establishing this conclusion except insofar as they reduce to standard skeptical doubts concerning scientific procedure and hence are not relevant here. And his account leaves as a complete mystery our practice of assigning rules to Robinson Crusoe, an exotic instance of what is in fact a standard case, as noted.

We should stress again that there are differences among these examples. The structure of a molecule of water or benzene determines how either substance will behave in a chemical experiment, but the structure of Jones's language does not determine what he will say, although it does determine (fairly closely) how he understands what he hears; and the structure of his initial state S_0 determines (again, fairly closely, perhaps very closely) what his language will be, given experience. Our theory of Jones—a grammar G of his language, a theory UG of the initial state of his language faculty—is a descriptive theory of his attained or initial competence, his knowledge system and its origins. Our theory is not a causal or dispositional theory concerning his behavior.

The rules of Jones's language are sometimes said to have "prescriptive force," but the term should be used with caution. These rules are not like the normative rules of ethics, for example. They entail nothing about what Jones ought to do (perhaps he should not observe the rules for one reason or another; they would still be his rules). And the question of the norm in some community is irrelevant for reasons already discussed. But whatever we conclude about the status of the rules, our theories about them are descriptive. We have every right to assert (tentatively, these being empirical matters) that it is a fact about Jones that his language has the rules R, R', \dots : rules of pronunciation, rules of binding theory determining the antecedent for a pronoun, rules determining the meaning of "John is too

stubborn to talk to" by a computational process, and so forth. It is a fact about Jones that with certain experiences, he understands *livid* to mean "flushed," and that with other experiences he would understand it to mean "pale." These are facts about Jones and his properties; in the case of S_0 , they are facts about the category of persons to which we assign Jones as we proceed to determine his specific properties. These properties of Jones enter into his behavior and understanding but do not determine his behavior, not even his inclinations or dispositions. But throughout, this is discourse about facts, facts about Jones. We can learn a good deal about such facts and already have done so.

Here, I think, an observation of Paul Horwich's is to the point. We may take Wittgenstein's skeptic to have undermined the "naive" claim

that there *must* exist inner states of understanding which explain our behavior. But that is not to question that there could be good a posteriori grounds for accepting such a picture. Nor is it to deny that, in either case, facts about meaning are as fully-fledged as facts ever manage to be (Horwich, 1984).

The same observation holds for rule following over a broader domain.

This seems to me correct, although I should add again that reference to a community seems to add nothing substantive to the discussion except under something like the Cartesian interpretation, moving to the level at which "form of life" corresponds to UG, that is, to attribution of "personhood"—a move that does not impugn the private model of rule following (and that is probably too narrow in its restriction to personhood as a matter of descriptive fact concerning normal usage and possibly potential science).

As persons, we attribute rule following to Jones on the basis of what we observe about him, perhaps without having reasons that justify this move. One might, but need not, adopt the intellectualized Cartesian account in this case. As scientists, we do attempt to develop a rather complex account in terms of postulated mental/physical states with elements that are not generally accessible to conscious introspection. The classical

problems about rule following remain unanswered; we have nothing like a "causal" account of behavior or any reason to believe that one exists. It also may be that the best theory will depart from the model of rule following altogether in any standard sense of the notion of rule, both for the receptive and the productive side of linguistic behavior, perhaps on the grounds already discussed, perhaps others.¹⁰

4.2. ON ATTRIBUTION OF RULE SYSTEMS AND RULE FOLLOWING

The discussion of the preceding section assumed the propriety of commonsense talk about rule following. This discussion accepted the view of rule systems of the sort rejected under the second conceptual shift discussed earlier, although in fact it could have been recast in terms of a principles-and-parameters model. Under this reformulation, we would not say, as scientists, that a person follows the rule of phrase structure (1) and the rules of passive and question-formation to yield (2):

VP → V NP C (1)

who was persuaded to like them (2)

Rather, the person uses the lexical properties of *persuade* under the projection principle, and the principles of Case adjacency, Move- α , binding theory, and so on, with values of the parameters fixed in a particular way. In earlier chapters, we regarded the question of choice among these and other theories as a matter of fact, and the skeptical considerations just reviewed do not seem to impugn this practice in any way.

Let us now put aside commonsense usage and the question of how, in ordinary life, we conclude that Jones is a rule follower who observes such-and-such rules. We are concerned now with the sense and legitimacy of the technical usage of our earlier discussion. Suppose that we find that the best theory we can construct attributes to Jones a certain mental structure S_L incorporating the I-language that includes the rule R and explains his behavior in terms of this attribution. Are we then

entitled to say that Jones follows R, or that his behavior is governed by R, as we have done so far?¹¹ Specifically, are we entitled to conclude from studies such as those outlined earlier that a person is following rules—say, those arrived at in the earlier discussion—when he assigns an interpretation to (2) or (3):

John is too stubborn to talk to (3)

It is of little interest to ask whether such principles as the projection principle or the binding theory conditions should be called “rules”; the latter term is too vague to allow an answer, and the answer would be of no significance in any event. We will refer to these principles, in UG and in their parametrized versions in particular grammars, as “rules,” but merely as a convenience; nothing turns on the matter. The more interesting question is whether we are entitled to hold that the person is in fact observing or somehow using the principles we are led to postulate in our study of the initial and attained state. Is behavior governed or guided by these “rules,” as we shall call them? Do the rules we postulate play what some call “a causal role” in behavior?¹² Do the principles formulated in UG concerning the initial state S_0 have “causal efficacy” in bringing about the attained state S_L ?

My assumption so far (as in earlier work) has been that we are entitled to propose that the rule R is a constituent element of Jones’s language (I-language) if the best theory we can construct dealing with all relevant evidence assigns R as a constituent element of the language abstracted from Jones’s attained state of knowledge. Furthermore, we are entitled to propose that Jones follows R in doing such-and-such (say, interpreting (2) and (3) in the manner he does) if, in this best theory, our account of his doing such-and-such invokes R as a constituent of his language. And if R is a constituent element of the initial state as determined by our best theory, and invoking R is part of our best account of why the attained state has such-and-such properties that then enter into behavior, we are entitled to propose that R has “causal efficacy” in producing these consequences. These assumptions have been regarded as highly controversial along with the underlying assumption

that psychological explanation, insofar as it involves rule following, is in principle part of the natural sciences.

Some analyses of rule following that appear to deny the legitimacy of this usage impose requirements so strong as to exclude the standard cases of commonsense usage as well; for example, Dennett’s proposal (1983) that we can speak of rule following only if there is a “very strong resemblance” to the case where we actually refer to an explicit representation of the rule in a formula (say, on a page, or as a “physically structured object, a *formula* or *string* or *tokening* . . .,” somewhere in the brain). As he seems to interpret “resemblance,” this proposal would rule out virtually all standard cases of rule following, linguistic or other, because they are not accompanied by recourse to such objects on a page or written in the brain. Or if something weaker is intended, it is entirely unclear what it might be or what bearing it might have on the cases that Dennett is discussing or that concern us (as he holds that his account does). Dennett’s further observation that rule systems might be only “tacitly represented” in the sense in which addition is represented in a hand calculator is correct but uninformative. Such possibilities cannot be excluded *a priori*. The question is one of best theory. It is also of no interest to observe, as Dennett and others do, that for every account that explains “input-output” relations in terms of rules, there is another with the same input-output properties that does not involve rules. We assume that this is so (for example, an account in terms of brain cells), but the question is whether in these terms we can formulate the applicable explanatory principles, provide explanations for a variety of facts within and across languages, or even state facts about words, phrases, repetitions, and so on—surely facts, if anything is. I will put aside objections of this nature, which are beside the point.

Many have taken exception to the idea, expressed by Jespersen and others since, that the speaker is “guided” by a perhaps unconscious “notion of structure” in forming free expressions or interpreting them. Quine, for example, has held that this is an “enigmatic doctrine,” perhaps pure “folly,” and that we may legitimately speak of “guiding” only when rules are consciously applied to “cause” behavior—surely not what happens in the ordinary use of language. Otherwise we may

speak only of behavior as "fitting" one or another system of rules: "Bodies obey, in this sense, the law of falling bodies, and English speakers obey, in this sense, any and all of the extensionally equivalent systems of grammar that demarcate the right totality of well-formed English sentences," the E-language (Quine, 1972); and we must refrain from inputting "psychological reality" to one or another system to which behavior conforms.

It is folly, then, to claim that one of a set of "extensionally equivalent systems of grammar" is correctly attributed to the speaker-hearer as a property that is physically encoded in some manner, whereas some other one merely happens to fit the speaker's behavior but does not correctly represent his knowledge. And it would be further folly to seek evidence that would distinguish two proposed grammars that generate the same sentences while attributing very different structural properties to them: say, a grammar that analyzed "John hit Bill" as a three-word sentence and another that analyzed it as a two-word sentence, the words being *Johnhi* and *tBill*; or a grammar that assigned to the sentence (4) the phrase structure indicated by brackets, as contrasted with a grammar that assigned it the phrase structure (5):

[his father][convinced Bill [that he should go to [a good college]]] (4)

[his][father convinced][Bill that he][should go to a good][college] (5)

Our explanations of the facts concerning the use and understanding of such examples as (2) and (3), however successful they might be, do not bear on any facts about the language faculty or even on its existence. There is no truth of the matter in such cases as those we have been discussing throughout, if we take Quine's position literally, just as there is no sense to the question of which of two grammars that generate the well-formed sentences of arithmetic in some notation is the "true" grammar.

Similar attitudes are revealed in certain approaches to the theory of meaning inspired by the work of Donald Davidson.

Michael Dummett describes Davidson's approach as holding that "the proper method" for the study of meaning

is to ask, for any given language, what body of knowledge would be required for someone to be able, in virtue of his explicit possession of that knowledge, to speak and understand the language. Here it is not maintained that any actual speaker really has such a body of knowledge, however tacitly or implicitly.

What the speaker does fits the theory in Quine's sense, but we must not go on to say that the speaker actually "has" the body of knowledge expressed in the theory. Dummett concedes that this way of "giving an illuminating account of what is involved in speaking and understanding the language" is "somewhat roundabout unless ability to speak a language actually does involve having such knowledge." He appears to be saying that it is somehow illegitimate to attribute possession of such knowledge to the speaker, even though knowledge of language may be "a genuine instance of knowledge." Something is lacking, some kind of relevant evidence, to prevent the account from being illegitimate, either unconfirmed or perhaps vacuous in principle.¹³

Many others have argued in a similar vein. John Searle, for example, holds that even if the rules proposed as elements of the attained state S_L explain the facts concerning such examples as (2) and (3), and the theory UG of S_0 explains why the attained state is of this form, thus providing a still deeper explanation for these facts, still this is not enough:

Additional evidence is required to show that they are rules that the agent is actually following, and not mere hypotheses or generalizations that correctly describe his behavior. It is not enough to get rules that have the right predictive powers; there must be some independent reason for supposing that the rules are functioning causally.

Searle holds that I have provided no evidence that the rules are more than "mere hypotheses" describing behavior.

Note that the rules proposed are not hypotheses describing behavior, for reasons already discussed, although they are of

course "mere hypotheses"—that is, theories of grammar and UG are empirical theories, not part of mathematics. Furthermore, it is not in doubt that evidence has been provided, however one may assess its strength, that the initial and attained states have the structure postulated in these attempts at explanation. So the objection must be that evidence that certain rules are constituent elements of the state attained does not bear on the question of whether the agent is following these rules, whether they are "functioning causally"; something more is needed. And evidence that certain assumptions about the constituent elements of the initial state offer a deeper explanation for facts about the agent's behavior does not in any way support the conclusion that these elements have "causal efficacy" in bringing about the state containing rules that the agent is in fact following.¹⁴ The objection must be, then, that even if there is evidence that S_0 includes the principle P and Jones's attained state S_L includes the rule R , and even if these conclusions are specifically and crucially invoked in the best account of Jones's behavior, still none of this would provide any reason to believe that R is a rule that Jones is following or that P has "causal efficacy."

I think that the qualms expressed by these and many other commentators are misplaced. To see why, let us compare three kinds of fact:

- (i) When Jones is dropped, he falls
 - (ii) Jones understands (2) and (3) in the manner indicated
 - (iii) Jones has acquired the knowledge that the facts are as they are on the basis of a certain exposure to data
- (6)

How do we proceed to explain these facts?

In each case, we proceed by attributing to the person certain properties: in case (6i), mass; in case (6ii), an attained state S_L incorporating a language (an I-language); in case (6iii), an initial state S_0 . Returning to Quine's example, Jones obeys the law of falling bodies because of some property that he has: mass. This property does not suffice to explain why he obeys the "laws of grammar," so we therefore have to attribute to him other properties, those of S_0 and S_L , just as we attribute properties

beyond mass to a molecule of water or benzene or to a cell of the body to explain their behavior.

Quine's comparison, which is common in the literature, is completely beside the point. A more accurate rendition of his statement quoted above would be the following: English speakers obey any and all of the extensionally equivalent systems of grammar that demarcate the right totality of well-formed English sentences in the sense in which bodies obey any and all of the extensionally equivalent theories that demarcate some set of phenomena projected in some undetermined manner from some arbitrarily selected proper subpart of the evidence relevant to physics. This terminological proposal, for that is all that it is, is of no interest to physics or to linguistics.

What is the nature of the states we attribute to Jones? As already discussed, we want to develop the best theory of Jones's behavior, and we find that this best theory attributes to Jones a language faculty with these states, which we can then attempt to characterize at various levels: in terms of neural elements or in terms of general properties of these elements at a certain level of abstraction. At the second level, we propose certain principles, parameters, representations, modes of computation, and so forth, and we seek to explain the facts in these terms, taking this account tentatively to express the truth about the language faculty. Although there are distinctions (see p. 241), we do the same when we attribute a certain structure to a molecule of water or benzene, to a cell of the body, to the sun to account for light emissions, or to a machine of some sort. If our best theory accounts for Jones's behavior by invoking these rules and other elements, we conclude that they enter into Jones's behavior and guide it, that they play a "causal role" in the sense of this discussion. Let us consider more closely whether these moves are legitimate.

The argument at issue has two steps: The first step involves the tentative conclusion that the statements of the best theory of the language faculty are true; the second, that the elements (rules, etc.) invoked to explain Jones's behavior in the best theory we can construct in fact guide his behavior. Let us consider these steps in turn.

The first step seems unproblematic: We try to construct the best theory of all the evidence and tentatively accept its statements as true. Two theories of the state attained S_L (two gram-

mars of the I-language abstracted from this state) might yield the same judgments of grammaticality or form-meaning correspondence (or any other subset of relevant facts), but yet differ in that one is a better theory and/or accords better with other evidence, so that one will be chosen over the other as factually correct. There are innumerable ways in which this might happen, and we cannot offer strict "criteria" to determine such a choice any more than we can anywhere in rational inquiry. Theory G of S_L might conform to some theory UG of the initial state that is verified for other languages too, while theory G' might not—a familiar case. Or theory G' might contain superfluous rules to specify facts determined by a proper subpart of G' that is essentially equivalent to G , as, for example, in the case of a theory G' that provides explicit phrase structure rules to bar *"who did John see Bill?" although in fact the conclusion follows from other parts of anyone's theory (see p. 99). Or we might find relevant evidence from the brain sciences to select between G and G' . In short, we are trying to discover the truth about the language faculty, opportunistically using any kind of evidence that we can find, and relying on the vague principles used throughout rational inquiry to find more elegant, deeper, and more empirically adequate theories.

Evidently, we will try to choose among "extensionally equivalent" theories of the state attained, meaning theories that coincide on some subpart of the evidence (e.g., grammaticality judgments, form-meaning correspondences, or whatever) or that coincide on "all the evidence" but differ in depth, insightfulness, redundancy, and other characteristics. This is just standard scientific practice. There is no general reason to doubt that these efforts deal with questions of fact; and apart from empirical uncertainties, there is no reason to hesitate to regard their conclusions as (tentatively) true of the language faculty. As Tyler Burge (1984) observes:

... questions of ontology, reduction, and causation generally, are epistemically posterior to questions about the success of explanatory and descriptive practices. One cannot reasonably criticize a purported explanatory or descriptive practice primarily by appeal to some prior conception of what a "good entity" is, or of what individuation or reference should be like, or of what the overall structure of science (or

knowledge) should turn out to look like. Questions of what exists, how things are individuated, and what reduces to what, are questions that arise by reference to going explanatory and descriptive practices. By themselves, proposed answers to these questions cannot be used to judge an otherwise successful mode of explanation and description.

The assumption that there is a distinct language faculty of the sort discussed is a relatively "successful mode of explanation and description," indeed the only one known in general terms.

Questions of a similar sort were raised in nineteenth-century science. In an interesting review of the period, John Heilbron (1964) observes that

the practicing physicist who was also a philosopher of science was (and perhaps still is?) something of a split personality. In the former capacity, the scientist might reason, write and speak as if his concepts were more real than [those of] his colleagues; in the latter role, somewhat self-conscious, perhaps, in the presence of those colleagues, he professes that his concepts have nothing more to recommend them than the convenience they offer in the arrangement of his ideas. In a study as rarified as the mathematical physics of the end of the last century it was probably impossible for most people to make progress without committing themselves so completely to their ideas that the philosophical line between the convenient concept and the ultimate reality vanished in practice. Yet at the end of the nineteenth century, one could not avoid the influence of Kant, with his insistence that the world of the noumena is forever closed to us; nor that of Comte and Mill, with their assurances that the aim of "true" science, positive science, is simply accurate description of the phenomenological world.

Thus on the one hand, Boltzmann described his molecular theory of gases as nothing but a convenient analogy, and Poincaré held that we have no reason for holding to a belief in the existence of matter apart from its convenience for explaining phenomena, and no reason to choose between ethereal-mechanical or electromagnetic theories of light; we accept the molecular theory of gases, he held, only because we are familiar with

the game of billiards. Yet on the other hand, Heilbron continues, "One has the impression from the way the concepts atom, ion, molecule and ether were used by scientists in everyday problems that many practicing physicists and chemists considered these particles as real as vacuum pumps and spectrosopes." Skepticism with regard to the realist stance diminished in practice as evidence converged (e.g., different ways of calculating the number of molecules in a volume of gas, results which, according to Lorentz, made it "no longer possible to doubt reasonably the real existence of molecules and atoms, i.e., of very small particles of matter separated one from another") and as more successful theories were advanced (e.g., Bohr's theory of the atom).

With all the many distinctions that should be drawn, the study of mind—language in particular—is in a somewhat similar state today. We proceed in practice by taking a realist stance toward theoretical discourse. Although many feel that this stance is somehow illegitimate, that it goes beyond the evidence, the issues are not crucially different from those that arise in any intellectual work of significance. The significant questions have to do with the persuasiveness and explanatory power of the theories and the quality and scale of the evidence bearing on them. No issue of principle arises in connection with the first step of the argument.

Let us turn now to the second step of the argument. Suppose that our most successful mode of explanation and description attributes to Jones an initial and attained state including certain rules (principles with parameters fixed or rules of other sorts) and explains Jones's behavior in these terms; that is, the rules form a central part of the best account of his use and understanding of language and are directly and crucially invoked in explaining it in the best theory we can devise. Are we now entitled to say that the rules of the attained state guide Jones's understanding and enter into his behavior, and that the rules of the initial state have "causal efficacy" in bringing about the attained state? Are these *further* claims, beyond the claim that our best theory of Jones attributes to him encoded rules that operate in this way? Since the first step of the argument exhausted the methods of science, if something else is involved in this second step, then it goes beyond the natural sciences; and

indeed it is sometimes argued that psychological explanation is unlike explanation in the natural sciences and cannot be assimilated to the natural science model because of its crucial reference to the "causal efficacy" of the rules that guide behavior.

I cannot see that anything is involved in attributing causal efficacy to rules beyond the claim that these rules are constituent elements of the states postulated in an explanatory theory of behavior and enter into our best account of this behavior.¹⁵ We will say that our theories of S_0 and S_L involve encoded rules that guide Jones's behavior when our best theories attribute these rules to Jones and resort to them in accounting for his behavior: an unsurprising answer, but one that cannot be improved upon. Clearly, we cannot obtain more evidence than all the evidence, or find better theories than the best theory. Nor is there any hope of identifying a kind of magical evidence that plays some unique role in determining that the rules attributed to Jones and invoked to explain his behavior in the best theory of all the evidence in fact guide Jones's behavior.

Returning to Searle's critique, his position is that to show that a person is following the rule R, one must show that the content of the rule plays a causal role in the production of the rule-governed behavior. What kind of evidence could confirm this conclusion, beyond best theory considerations of the sort just indicated? One possibility is that nothing more is required, in which case recourse to this condition is mere incantation. A second possibility is that more evidence is required than all the evidence and a better theory is required than the best theory; this we may dismiss. The only other possibility is that some special kind of evidence is required to show that the computational principles that constitute part of Jones's knowledge and are involved in his use of language (and hence have "psychological reality," if that hopelessly misleading and pointless term is invoked) are indeed rules that Jones is following. What kind of evidence would this be? Some have held that access to consciousness would count as such evidence, but this is plainly wrong, as already noted; this is generally evidence of the least useful and relevant sort, far weaker than evidence of the kinds adduced earlier. Other kinds of evidence that one can imagine (neurophysiological, etc.) might well be valuable, but they have no magical status beyond that of the evidence regularly

adduced to establish theories of knowledge and behavior. In fact, the quest seems entirely misguided, whether at the level of grammar or UG.

Consider a specific example (borrowed from Searle). Let us say that Jones follows the rule R: Drive on the right-hand side of the road. Consider now the rule R': Stay on the side which is such that the steering wheel of the car is nearest to the center line of the road. Suppose we want to determine whether Jones is following R or R', both of which correctly describe ordinary behavior. How then would we proceed?

We might ask Jones what rule he is following, but this possibility arises only because of the triviality of the example and is not available in the general case, for example, the case of following a rule of pronunciation or syntax. Furthermore, such evidence is at best very weak because people's judgments as to why they do what they do are rarely informative or trustworthy.¹⁶ So other approaches are necessary.

We would try to construct other cases that would distinguish the two rules. For simplicity, assume that Jones consistently follows his rule, either R or R'. An obvious test would be to try Jones in a British-made car with the steering wheel on the right. Suppose we discover that Jones follows R and violates R', which must then be revised to R'': Stay on the side which is such that the steering wheel of the car is nearest to (farthest from) the center line of the road if the steering wheel is on the left (respectively, right). In the context of normal science, we would presumably terminate the inquiry right here, but let us proceed.

Suppose we test Jones on a multilane highway, and we find that he veers right (i.e., away from the center line) to avoid a head-on collision with a car coming toward him on the right-hand side. Then we have evidence that he follows R, not R'', and we must somehow modify R' to R'''. Or we may try the same experiment on some area with no demarcated center line (say, a beach) and find the same results, again supporting R over R''. In principle, we could seek neurological evidence. Suppose that some drug X affects the brain in such a way as to eliminate the notion "right" while leaving whatever concepts appear in R', whereas some drug Y has the opposite effects. Suppose we find that after taking X Jones loses the capacity in

question while Y does not affect the capacity. Then we have evidence for R over R'. One can imagine many kinds of evidence; it is simply a matter of ingenuity and feasibility. This is exactly how we would proceed, in the general case, to determine whether it is R or R' that Jones is following, and to determine which rule is "causally efficacious" in his behavior.

In short, we try to find the best theory of all the evidence, and if that theory invokes R to explain Jones's behavior, we say that Jones is following R.

Suppose that our efforts to distinguish R and R' are in vain. To turn to a language case, suppose that R and R' are two rules proposed to account for the facts concerning (2) or (3), and the two yield equally good theories of all the evidence. We now face a true indeterminacy, not of the uninteresting sort that we know always exists in an empirical discipline, but of a real and significant sort. How do we proceed at this point?

We ask: What is it about these theories that makes them work? Is there some deeper level at which the theories are really identical, the latter being the right level for our theory of language? We ask whether the two theories are overly concrete realizations of the same system of principles that should be formulated at this more abstract level—more or less in the way that rotations of a plane figure and elementary operations of arithmetic are two realizations of the principles of group theory. We try to find a more abstract principle P incorporating just the relevant features shared by R and R' without the extra structure that differentiates these rules. If we succeed, we then conclude that this deeper principle P is the actual rule that John is following. This is the right way to deal with significant indeterminacies, which are not some lethal defect or source of paradox, as assumed in much current philosophical discussion, but rather constitute important evidence that can guide our thinking in theory formation.

Incidentally, we should expect something of the sort just described in actual practice; we should expect that the actual theories we work with for concreteness are "too concrete," that is, involve elements that are really playing no explanatory role and that we should try to eliminate as we seek to discover what is really doing the work of explanation. It is a difficult but important task to compare successful theories in this manner

and to use this analysis to progress toward the correct theory of the language faculty.¹⁷

So far, this is standard scientific practice. We say that Jones is following R if R is attributed to Jones's language (or its initial state) at the relevant level of abstraction in the best theory of all the evidence and the best account of Jones's behavior invokes the rule R, just as attribution of mass to Jones is invoked in the best theory explaining how he falls. But the argument we are considering holds that this does not satisfy the requirements for establishing that Jones is following R. We have to show, further, that the form or semantic content of R plays a "causal role" in Jones's behavior. But this makes no sense at all.

As already observed, it makes little sense to try to identify specific kinds of evidence or specific criteria that play just the role of identifying R as a rule that Jones follows.¹⁸ To return to the analogy discussed before (with provisos indicated), suppose that we are nineteenth-century chemists postulating valences, benzene rings, properties expressed in the periodic table, and so forth. Suppose someone were to ask what justification we have for doing so and we were to say: This is the best theory we can devise to deal with all the evidence we have. Would it be reasonable to demand more? For example, to demand specific criteria to determine whether there are elements with a valence of 2? Would it have been reasonable to ask nineteenth-century chemists to state explicitly the conditions that would justify their saying that the entities they postulate are "represented" in physical mechanisms?

All of this seems senseless. In no domain is there any hope of satisfying demands of this kind. The chemists could do no more than observe that the entities they postulate form part of the best theory of all the evidence, and that they would like to learn what physical mechanisms of some different order account for these facts. To say that Jones follows rule R is to say no more than that the best account of Jones's behavior invokes the rule R of Jones's attained state (and uses the mode of computation postulated, etc.). We cannot go beyond the best theory of all the evidence we have. There is no reason to depart from the usual procedure of the sciences, taking a realist stance with regard to theoretical terms. There is no reason to suppose that some

crucial evidence is missing when we conclude that Jones is following R on the grounds that the best theory accounting for Jones's behavior includes R as an operative rule, entering into explanations of this behavior. As Demopoulos and Matthews (1983) observe, "the apparent theoretical indispensability of appeals to grammatically characterized internal states in the explanation of linguistic behavior is surely the best sort of reason for attributing to these states [and, we may add, to their relevant constituent elements] a causal role in the production of behavior."

We do, of course, assume that the rules are somehow represented in physical mechanisms, and, like the nineteenth-century chemist, we look forward to the day when something will be learned about exactly how this takes place. But there is no point in a demand that we give answers to questions that no one yet understands and concerning which no one yet has any idea what would be relevant evidence.

Turning back to some of the objections mentioned at the beginning of this section, recall that Dummett asks us to seek a formulation of the body of knowledge that would be required to speak and understand "any given language." But what does it mean to refer to "a given language"? An infinite class of expressions, or expressions paired with meaning or use conditions, or actions, or whatever, is never "given." What is "given" is some finite object, a finite set of observed phenomena. On the basis of observed phenomena, the child's mind constructs a language (an I-language); the child comes to know a language. The linguist attempts to discover the nature of this process of coming to know a language, and to determine what body of knowledge has been acquired, thus affording the ability to speak and understand. The linguist's grammars and theory of UG are theoretical proposals as to what is known and how it comes to be known. They offer the only possible kind of answer (although in practice, no doubt, the wrong answers in part at least) to Dummett's demand.

What about Dummett's apparent belief that some kind of relevant evidence is lacking that would legitimize the assumption that a particular theory concerning the knowledge attained is the correct one, the correct characterization of what the speaker-hearer knows? This seems to me seriously in error.

True, we always want more evidence, and evidence of more varied types, but there is no defect of principle in the kinds of evidence we have, no defect of principle that other kinds of evidence would remedy. In fact, the kinds of evidence we have carry us rather far toward determining the properties of the system that constitutes "knowledge of a given language." We naturally will seek evidence that carries us even further, but the study of language and meaning is hardly unique in this regard. In fact, the qualms that Dummett attributes to the Davidsonian view and that he appears to share are no more relevant here than in any other empirical inquiry.

Dummett's argument that we are not entitled to impute unconscious knowledge of the rules of language appears to be based on the belief that this is somehow circular or vacuous, like the pseudoexplanation that a band stretches when pulled *because* it is elastic (recall that he does regard knowledge of language as a "genuine instance of knowledge," although by "language" here he presumably means some version of E-language). He alleges further that I "never expressly [adduce] evidence that our linguistic competence is based on unconscious knowledge at all, rather than being a skill," an ability to speak and understand based on reflexes or feedback mechanisms or whatever. Understanding may be "simply a practical ability—say, to respond appropriately," and we therefore achieve nothing by attributing unconscious knowledge of the rules of language. "The difficulty is, however, that we have no idea what structure and character knowledge, conceived as an internal state, may have, apart from the structure of what is known." We identify knowledge "solely by its manifestations." "We therefore need much more than Chomsky offers us about how, in general, unconscious knowledge is manifested if we are to understand or evaluate his theory." We need some kind of independent check, beyond the structure of what is known, to make the explanation nonvacuous. Let us consider these objections in turn.

It cannot be that our account, say, of what is involved in understanding (2) or (3) is circular and vacuous, since it clearly could be wrong and we can easily produce hypothetical evidence that would disconfirm it; as discussed earlier, it is based on principles that have a broad range of empirical consequences.

Hence the account is in no way similar to the pseudoexplanation of the stretching of a band. It is simply false that no evidence has been adduced that our linguistic abilities are based on rules and principles that are somehow represented in the mind (as to whether this should be called "knowledge," see the next section); certainly evidence has been presented, however one estimates its force, and many other kinds of evidence that would be relevant if obtainable have been sketched. Someone who believes that this kind of understanding is "simply a practical ability," a "skill" of some sort, owes it to us to provide some account of the nature of this ability or skill; otherwise, the proposal is empty (see also pp. 9–13). Consider the statement that we have no idea what the structure and character of an internal state of knowledge may be apart from the structure of what is known. Dummett is unclear about what he thinks "is known," but he seems to mean that what is known is that sentences mean such-and-such, the elements of E-language in some version. But what, then, is "*the structure* of what is known"? This is hopelessly obscure. On the other hand, we have quite a good idea of the structure and character of (possession of) knowledge as an internal state, and we have nontrivial theories concerning the matter; these theories are based on evidence as to what "is known" in Dummett's sense, and in principle much else, and are certainly not vacuous or circular. To say that we identify knowledge (or the structure of knowledge, or the internal state of knowledge, or the system of rules constituting knowledge, etc.) "solely by its manifestations" is true only in the sense that the nineteenth-century chemist identified the structure of benzene "solely by its manifestations." In fact, we identify the system of knowledge of language that accounts for facts concerning (2), (3), and so forth by such manifestations of this knowledge as the judgments concerning referential dependence, by judgments concerning other expressions, by behavior of speakers of other languages, and in principle in many other ways as discussed earlier. We have quite extensive accounts of how not only in general, but also in many specific cases, "unconscious knowledge is manifested," and if these do not serve to indicate how we are to understand or to evaluate the theory that incorporates these cases, Dummett owes us a reason as to why they do not and as to how the situation is different, in

principle, from familiar cases in the empirical sciences. Dummett's objections amount to saying that we confirm a theory by determining how it accounts for available evidence, and therefore we have no basis for affirming the principles of the theory.

Perhaps the fear is that the evidence is all "of the same sort," primarily informant judgments, and that other types of evidence are necessary. As an objection of principle, this is plainly without merit; these phenomena certainly constitute evidence, and in fact the evidence they provide does suffice to confirm or to refute proposed theories and even leads to empirical theories of some scope and depth. As an objection of a narrower sort, one can take it seriously as an argument that the evidential base is too narrow to carry conviction; one who believes this might ask what other kinds of evidence would strengthen or undermine the theories we are led to construct on the basis of the (not inconsiderable) evidence that we can now readily obtain. In practice, what has been proposed along these lines has not been very informative, but certainly any improvement in this regard will be welcome.¹⁹

As for Quine's conclusion, he assumes that the possibilities are exhausted when we speak of behavior as "guided" by conscious rules that "cause" it or as simply fitting rules we state. But there is no reason to accept this conclusion. Behavior is guided by the rules and principles of a system of knowledge, and these are, in fact, generally not accessible to conscious awareness. The conclusion is perfectly intelligible and in fact is the only one yet formulated that appears to be at all warranted by the known facts. On these assumptions, we can explain a good deal about the ways in which linguistic expressions are used and understood, although we cannot predict what people will say. Our behavior is not "caused" by our knowledge, or by the rules and principles that constitute it. In fact, we do not know how our behavior is caused, or whether it is proper to think of it as caused at all, but that is another matter entirely. As for Quine's tacit assumption that the notion of E-language or of grammars extensionally equivalent over E-language is relatively unproblematic as compared with I-languages and their grammars, that is incorrect for reasons already discussed (see pp. 26–31).

In general, there seems to be no force to a wide range of objections of this nature in the philosophical literature, many of which I have discussed elsewhere.²⁰

There are, to be sure, further distinctions to be made with regard to how the rules of language (the principles of S_0 , etc.) are used. We mentioned earlier the classical view that the use of these rules in speech is free and indeterminate. To rephrase the point in the terms of contemporary discussion: the cognitive system involved in use of language is "cognitively penetrable" in the sense of Pylyshyn (1984) and other current work; that is, our goals, beliefs, expectations, and so forth clearly enter into our decision to use the rules in one way or another, and principles of rational inference and the like may also play a role in these decisions. This is true not only of what we decide to say but of how we decide to say it, and similar factors enter at some level into determining how we understand what we hear.

Consider, for example, the pronunciation rule P, discussed earlier, that merges lax and tense /i/ before /g/ in dialects in which *regal* and *wriggle* are pronounced alike. A speaker of this dialect may choose to ignore the rule, and to apply instead the rule of the standard dialect, for one reason or another; this may be done, in fact, without any awareness of what the rule is or even that it exists. Or consider again the "garden path" sentence (7) (see Chapter 2, note 12):

the horse raced past the barn fell (7)

In normal circumstances, a speaker of English presented with (7) will judge that it is not a sentence; this decision is presumably reached by applying the rules R_1 that interpret "the horse raced past the barn" as a clause, so that *fell* has no role and the principle FI (full interpretation) is violated. But with prior instruction or experience, the same speaker might well use the different rules R_2 , interpreting "the horse raced past the barn" as a reduced relative clause construction, the subject of "fell," so that (7) is given the meaning "the horse that was raced past the barn fell." It is entirely possible that quite different rules are used in the two cases, so that the process of sentence perception

is cognitively penetrable, influenced by factors that might involve expectations, reasoning, and so on.

But while the system of language use is cognitively penetrable in this sense, the system of principles of S_0 presumably is not; it merely functions as a kind of automatism. In normal cases, the same is true of the system of speech perception and much of speech production; it may well be true, for example, that the rules R_3 used in parsing such sentences as (3), repeated here, form a cognitively impenetrable set:

John is too stubborn to talk to (8)

There is a distinction to be made between cognitively impenetrable systems that constitute what Pylyshyn (1984) calls "functional architecture" and systems that involve reference to goals, beliefs, and so forth, and perhaps inference of one sort or another. In Pylyshyn's terms, the distinction is between the "symbolic (or syntactic) level" and the "semantic (or intentional) level," each to be distinguished from a third level, the "biological (or physical) level" at which description and explanation are in terms of laws of physics, biochemistry, and so forth.

In these terms, most of our discussion so far has been at the "symbolic level," not the "semantic-intentional level." It is at the symbolic level that we have proposed a system of rules, representations, and computations to account for the facts. Notice that it would be rather odd to argue that rules and representations enter into the cognitively penetrable system of language use—e.g., the rule of pronunciation P or the rules R_1 , R_2 , and FI that enter into the interpretation of (7)—but that these very same entities are not rules and representations when they constitute elements of the symbolic level; or that the rules R_1 , R_2 , and FI exist at the symbolic level only if the system using them in parsing is cognitively penetrable as suggested, but not, say, if parsing actually provides both interpretations of (7) (namely, by using R_1 , R_2 , and FI , if that is what the best theory asserts) and some other system selects between them; or that we are not entitled to speak of rules and representations at the symbolic level when we provide an explanation for the interpretation of (8) and similar cases in terms of R_3 and the representations involved; or that we cannot speak of the prin-

ciples of S_0 in these terms, even though these principles (or their parametrized variants) are elements of the cognitively penetrable system of language use, at the semantic-intentional level. Rather, it seems that at each level we are entitled to postulate rules and representations, and to hold that these are involved in language use, when "best theory" considerations of the sort discussed lead to this conclusion.²¹

4.3. ON KNOWLEDGE OF RULES

Much of the interest of the study of language, in my opinion, lies in the fact that it offers an approach to the classical problem that I called "Plato's problem": the problem of explaining how we can know what we do know. Plato's answer was that much of what we know is inborn, "remembered" from an earlier existence. Leibniz argued that the idea is basically correct, but must be "purged of the error of preexistence." Much of our knowledge is innate, he held, virtually present in the mind even if not clearly articulated. This is true of the propositions of arithmetic and geometry, and also of "the in-built principles of the sciences," and of practical knowledge. Within a different philosophical tradition, David Hume spoke of those parts of our knowledge that are derived "from the original hand of nature," echoing Lord Herbert's reference to "that part of knowledge with which we were endowed in the primeval plan of Nature." Rather like Leibniz, Hume regarded such innate knowledge as "a species of instinct."

These ideas are, I think, basically correct. We can now just begin to flesh them out in various domains. Suppose we are presented with a plane figure perpendicular to the line of vision, rotating until it disappears. Suppose that under a certain range of conditions, we take it to be a rotating plane figure, not a plane figure shrinking to a line. Given a series of visual presentations, we perceive a rigid object moving through space, not an object changing its shape. The judgments we dismiss are consistent with the evidence presented, but they are rejected by our system of interpretation of objects in visual space. If the judgments we make are correct, then we have true knowledge, propositional knowledge that the plane figure is rotating, that

a cube is moving through space, and so forth. Similarly, we have propositional knowledge that sentences mean so-and-so.

How do we derive such knowledge? In the case of perceptual space, it seems that the visual system is designed to implement the tacit assumption that objects are rigid—Shimon Ullman's "rigidity principle" (1979). Roger Shepard (1982) argues that the perceptual system has, "over the eons of vertebrate evolution, internalized the most important invariants and constraints in the external world," including properties of rigidity and symmetry of objects. In the language case, it seems that our propositional knowledge derives from the interaction of principles of UG once parameters are fixed. What the evolutionary origins of these principles might be is unknown, although some vague analogies have been considered in terms of hierarchical properties, locality principles, and the like, and there are some suggestions about possible functional properties of efficient processing.²²

It is known that the character of the visual system can vary over a certain range depending on early visual experience, as the language faculty plainly can, yielding the diversity of languages. Learning is a matter of fixing the system within the permissible range; in the language case, by setting the parameters of UG and adding a periphery of marked exceptions. What we know is then determined by the functioning of the mature system, sometimes involving moderately complex inferencelike computations. Such knowledge that so-and-so is not in general warranted or justified by the presented evidence in any useful general sense of these terms, nor is it self-evident or "self-presented" in Roderick Chisholm's sense. An organism differently endowed, or with the same endowment but with a mature system fixed in a different way by early experience, might have a different range of knowledge, understanding, and belief, and might interpret presented experience differently. Furthermore, various cognitive systems appear to operate in quite different ways, as determined by our biological endowment. It may well be that something similar is true of Leibniz's "built-in principles of the sciences" and practical knowledge, although not in quite the sense he intended. If this is correct, then the paradigms of much traditional and modern epistemology are inadequate,

although we can give a plausible reinterpretation of certain classical ideas.

One might ask whether it is proper to use the ordinary language term "knowledge" in this connection. Is it, for example, proper to say that a person who knows a language in the ordinary sense "knows the rules of the language" (the I-language) in the technical sense? In part, the answer is certainly negative, because I-language, like other technical notions of scientific approaches, is not language in the pretheoretic sense, for reasons discussed earlier. It is not clear that much is at stake here; our intuitive concept of knowledge becomes hazy and perhaps misleading at certain crucial points, and ordinary usage in fact differs from language to language; one does not speak of "knowing a language" but rather of "speaking" or "understanding" it in languages very similar to English, although this does not affect our concern to discover the cognitive system—whether we call it "knowledge of language" or something else—that enters into our knowledge of particular facts, say, the facts concerning (2) and (3) of Section 4.2.

I think that for the theory of knowledge, we need a concept that is close to the term "know" where it is clear, but that may sharpen or extend its normal usage, much as in the case of the term "language" discussed earlier. Elsewhere, I have suggested that we make up a term, say "cognize," assigning it the following properties. When we know that *p*, we cognize that *p*. Therefore we cognize that the sentences (2) and (3) have the range of meaning they do have. Furthermore, we cognize the rules and principles from which these cases of knowledge-that derive, and we cognize the innately given principles that are further articulated by experience to yield the mature system of knowledge that we possess. The term "cognize" is similar to "know." It is identical with regard to knowledge of specific facts, and also, I think, virtually so with regard to knowledge of the system that underlies them, subject to the provisos noted concerning the difference between the term "language" of ordinary usage and our technical term "language" in the sense of I-language. Let us consider some of the relevant cases.

Knowledge of language involves (perhaps entails) standard examples of propositional knowledge: knowledge that in the

word *pin*, /p/ is aspirated, whereas in *spin* it is not; that the pronoun may be referentially dependent on *the men* in (9i), but not in the identical phrase in (9ii), and so forth:

- (i) I wonder who [the men expected to see them] (9)
 (ii) [the men expected to see them]

If these are not instances of knowledge, it is hard to see what is. In this case, the person who knows the language knows these facts, and cognizes them.

Suppose that R is a rule of English grammar that states that verbs cannot be separated from their objects by adverbs, so that in accordance with R, the sentence "I read often the newspaper on Sunday" is unacceptable; rather, we say "I often read the newspaper on Sunday." Suppose that John, a speaker of English, follows the rule, but Pierre, who is learning English, does not and regularly produces and accepts the sentences R marks unacceptable, as in his native French. What we would say, in this case, is that John knows that verbs cannot be separated from their objects by adverbs, but Pierre has not yet learned this and does not know it. Thus John knows that R, but Pierre does not know that R. Of course, we cannot assume that John knows that R holds, obtains, is a rule of his language. John quite probably does not know this, although some linguist may. In other words, there is no legitimate "semantic ascent" from "John knows that R" to "John knows that R holds."

Suppose, however, that R is not itself a rule of English but rather a consequence of the rule R' that states that Case assignment in English observes a strict adjacency requirement: The value for the Case assignment parameter in English is *strict adjacency*. This seems a plausible conclusion, as we have seen. Would it then be proper to say that John knows that R', but Pierre does not—that is, that John knows, but Pierre has not yet learned and does not know that the value for the Case assignment parameter in English is *strict adjacency*? I think many people would be reluctant to say this of John and Pierre, although I would not. The reason for the difference between attributing knowledge of R and R' does not lie in the nature of the state of knowledge of John or Pierre, but rather in the familiarity of the notions verb, adverb, and object, which enter into R, as com-

pared with the unfamiliarity of the notions Case assignment and adjacency parameter, which enter into R'. But this is irrelevant to the description of John's or Pierre's state of knowledge: These states are what they are, independently of our knowledge of linguistic theory.

In fact, there is nothing about the case that is specific to knowledge of language. Suppose that the sun emits light because of processes of fusion internal to the sun. The statement that this is so is not a meaningful statement of my language if I do not know the meaning of the word "fusion," unless we allow, as perhaps we should, that the meaning of the term for me is expressed in terms of the knowledge of others in my speech community (see p. 18). In this case, however, I think we should attribute knowledge that R' to John, but not Pierre, on the same grounds.

Assuming this to be correct, let us return to examples (9). Suppose our best theory asserts that speakers know the facts of referential dependence in these cases because their language provides the representations (10i) and (10ii) for (9i) and (9ii), respectively, with the interpretation discussed earlier, and contains the principles assumed earlier—in particular the principles of the binding theory and those determining the presence and character of empty categories in mental representations:

- (i) I wonder [who the men expected [*e* to see them]] (10)
 (ii) the men expected [PRO to see them]

As we have seen, there is good reason to suppose that something like this is correct. Should we then say that the person who "has" this language "knows the binding theory principles" and so forth? The case is very much like that of R' in the preceding example, and a positive answer seems consistent with normal usage.

If this reasoning is accepted, then "know" is very much like "cognize." If not, then "know" departs from "cognize" in this respect and is not an appropriate term for the theory of knowledge, because the correct account of the state of knowledge attained by John and Pierre should be independent of what we know; it is a statement about them, not us. I do not think that the question is a very important one. If we continue to use the

term "know" in the sense of "cognize," as seems reasonable, then it is proper to say that a person knows that R, where R is a rule of his or her grammar.

The same is true in many other cases. Let R now be the rule that stops in English are aspirated in initial position but not after /s/. If John observes this rule, I would have no hesitation in saying that he knows that R: and if Pierre does not, that Pierre has not yet learned and does not know that R, although he may come to know that R with further experience. Someone unfamiliar with their terms "stop" and "aspiration" would not be prepared to describe the state of knowledge attained by John and Pierre in these terms, but that is irrelevant to the correct attribution of knowledge.

Let us consider now some principle P of UG that is immutable and not parametrized, say, the principle that pronominals cannot c-command their antecedents, one of the principles that entered into the explanation of why the meaning of (11i) and (11ii) is what it is, and of how the facts are known without direct experience:

- (i) John is too stubborn to talk to Bill (11)
 (ii) John is too stubborn to talk to

Should we say that John knows that P, assuming now that we understand the terms that enter into the formulation of P?

My own rather vague intuition is that there is something strange about this. On the other hand, suppose that some Martian, who does not share the initial state S_0 of the language faculty, is being taught a human language and comes to learn that P. Then I think we would have no hesitancy in saying that before he did not know, and now he does know, that P. But his state of knowledge now could be the same as that of John, so that it would seem that we should say the same about John. The difference, if it is real, again seems independent of the actual states of knowledge of the individuals we are describing, and thus should be eliminated from a principled theory of knowledge. In this case, then, the term we need for the theory of knowledge appears to be something like "cognize," which may abstract from certain features of the term "know" of informal usage.

In most of these cases, there is no way for one to determine by introspection that the rules and principles hold. One cannot become aware that one knows, or cognizes, these rules and principles. If presented with these principles as part of a theory of grammar, we may become convinced that they are correct, but we do so "from the outside," as we may be convinced that a theory of fusion correctly explains the emission of light from the sun. Suppose that the facts were different, and that we could become conscious, by thought and introspection, that we do in fact make use of these rules and principles in our mental computations. Then, I think, one would have no hesitation in saying that we know them. If so, then cognizing would appear to have the properties of knowledge in the ordinary sense of the term, apart, perhaps, from accessibility to consciousness. We might say that "cognition" is unconscious or tacit or implicit knowledge. This seems to me correct, with the provisos given.

It has been argued that it is wrong, or even "outrageous,"²³ to say that a person knows the rules of grammar, even in the sense of tacit or implicit knowledge. As a general statement, this cannot be correct. We do not hesitate to say that John knows, whereas Pierre does not know, that verbs cannot be separated from their objects by adverbs or that stops are aspirated except after /s/—assuming, of course, that we know the meaning of the terms used in these ascriptions of knowledge. Recall that it would be wrong to say that John knows that the rule holds, but that is a different matter. Whether it is also proper to use the term "knowledge" in other cases such as those discussed seems to me unclear, because the concept is unclear, but in any event not very important for the reasons mentioned.

In these and many other cases, it seems that to speak of knowledge of rules, following rules, and so forth is reasonably in accord with normal usage, except, of course, with regard to the normative-teleological aspect of the commonsense notion of language. But again, not much seems to be at stake; different terms can be invented if one is uncomfortable with what appears to conform closely to normal usage (although not various philosophical theories) in these cases.²⁴

Let us return to Dummett's questions concerning the concept of unconscious knowledge and his belief that we need much more understanding about "how, in general, unconscious

knowledge is manifested if we are to understand or evaluate" the theory that knowledge of language is unconscious knowledge. We "need an account of how unconscious knowledge issues in conscious knowledge," and, he maintains, no such account has been given. But it seems reasonably clear, both in principle and in many specific cases, how unconscious knowledge issues in conscious knowledge, and the literature²⁵ contains many accounts of what the process might be. Thus, according to the theory that Dummett finds problematic or unintelligible, a person has unconscious knowledge of the principles of binding theory, and from these and others discussed, it follows by computations similar to straight deduction that in (9i) the pronoun *them* may be referentially dependent on *the men* whereas in (9ii) it may not, and that in (11i) the subject of *talk to* is *John* whereas in (11ii) it is not. That this is so is conscious knowledge, among the numerous consequences of principles of UG, which are surely not accessible to consciousness. It does not seem problematic to entertain the hypothesis that the mechanisms of mind permit something akin to deduction as part of their computational character. So we have a reasonably clear account in such cases as these of how unconscious knowledge issues in conscious knowledge.

We do not, of course, have a clear account, or any account at all, of why certain elements of our knowledge are accessible to consciousness whereas others are not, or of how knowledge, conscious or unconscious, is manifested in actual behavior. These questions are interesting and important, but not strictly relevant in this connection.²⁶ Whatever we take the internal state constituting knowledge of language to be—and surely there is such a state—these questions will arise.

Dummett holds, in particular, that knowledge of meaning is not susceptible to analysis in these terms, because

though it is conscious knowledge, [it] does not consist in our being able to *state* the meaning. We therefore stand in need of an account of conscious knowledge, and conscious thought, not carried by a linguistic vehicle. Until we can give such an account, we cannot say how much of our linguistic ability can be explained in terms of conscious but un verbalised knowledge; nor can we say how compelling

the case for its resting upon unconscious knowledge, or for there being such a thing at all, can be made to appear.

The passage seems to me rather obscure. I see no reason to believe that knowledge of meaning is conscious knowledge, or that it is in general accessible to consciousness. It is true that we know the meaning of the word "knowledge," for example, although we have great difficulty in determining and expressing this meaning. But in what sense is this conscious knowledge? Suppose it is true, as has been proposed, that the semantic analysis of words involves such fundamental conceptual elements as thematic relations including (abstract) position and its changes, agency, function, constituency, origin, and the like.²⁷ Do we have conscious (although un verbalized) knowledge of the way these elements enter into semantic analysis, if we know the meaning of the words? Or must we be able, in principle, to determine by introspection, or to recognize when informed, that these or other concepts are the correct ones? The answers are negative, throughout. It is true no doubt that there is such a thing as conscious but un verbalized knowledge: for example, our knowledge of properties of perceptual space and the behavior of objects in it. I do not see how this impugns the attempt to explain aspects of our linguistic abilities in terms of explicit theories of unconscious knowledge, which appear to have considerable explanatory force and to be unique in this respect. Dummett's assumptions as to the burden of proof seem arbitrary and unjustified.

The attempt to gain some insight into philosophical problems from the careful study of language has been a dominant theme of modern philosophy and undoubtedly a fruitful one. One question that has been relatively unexplored until recently is whether and how the scientific study of language might contribute to this end. One might ask whether there really is a "scientific study of language." My own view is that such a field is beginning to take shape. We can begin to see what the nature of such an inquiry should be and how it might become assimilated to the main body of the natural sciences. And there are some beginnings of explanatory theory that seem not inconsiderable. On the basis of these glimmerings of understanding, it seems that there may well be significant implications with

respect to certain classical problems of philosophy. My own guess is that these implications may prove to be richest with respect to the theory of human knowledge and understanding and more generally with respect to the nature of mind. In particular, what we are now coming to understand suggests that some of the questions of the theory of knowledge should be recast. Certain systems of knowledge that are central to human thought and action do not have the properties that have often been assumed to be paradigmatic.

Of these systems, language appears to be the one most accessible to study. There is reason to believe that knowledge of language, which provides an unbounded range of propositional knowledge and enters into complex practical knowledge, should be regarded as a system of principles that develops in the mind by the fixing of values for certain parameters on the basis of experience, yielding systems that appear to be highly diverse but that are fundamentally alike in deeper respects. The propositional knowledge that results is not grounded or warranted in any sense of these terms that will bear the burdens required in standard ways of thinking about these problems. There may be principles of learning, such as the subset principle (see p. 146), but it appears that the system of knowledge attained is largely preformed, as much a part of our biological endowment as is the general organization of our body. It seems that Plato's problem should be addressed along these lines.

We might speculate that the same is true in other areas where humans are capable of acquiring rich and highly articulated systems of knowledge under the triggering and shaping effect of experience, and it may well be that similar ideas are relevant for the investigation of how we acquire scientific knowledge in those domains where we are capable of doing so, because of our mental constitution. These systems, then, provide the frameworks within which our understanding can develop and flourish. The cost of this richness of potential in certain domains is the existence of limits in others, perhaps even absolute limits.²⁸ The relationship between scope and limits has traditionally been recognized but often thought to be a characteristic of the nonhuman animal world. It is the richness and specificity of instinct of animals that accounts for their remarkable achievements in some domains and lack of ability in

others, so the argument runs, whereas humans, lacking such articulated instinctual structure, are free to think, speak, discover, and understand without such limits (Herder). Both the logic of the problem and what we are now coming to understand suggest that this is not the correct way to identify the position of humans in the animal world.

In my opinion, it is within this larger framework that the technical developments that have taken place within the field of generative grammar should be understood. And is it this range of questions, still on the horizon, that gives them a broader significance that may prove to be far-reaching in the study of human nature and its specific manifestations.

NOTES

1. See Chomsky (1968, 1975b, 1980b).
2. See references of note 1. See also Chomsky (1966) and Bracken (1984).
3. I will discuss these only insofar as they bear on the production problem, without implying that a variant does not arise in the case of the perception problem.
4. Note that one must not assimilate Kripke's notion of "normative" in this discussion to the "normative-teleological" notion discussed in Chapter 2.
5. I have modified Kripke's citation insignificantly to accord with the text in Wittgenstein (1953).
6. What is now sometimes called the "Turing test," understood to provide evidence for possession of higher intelligence.
7. If this is the right approach; it may not be, as discussed earlier.
8. P. 36n. On the significance of these points, see Kirsh (1983).
9. See Gallistel (1980).
10. I am indebted to James Higginbotham, Alex George, and Jerry Fodor for comment on an earlier version of this and the following section.
11. Here and henceforth I will construe the term "behavior" broadly, to include interpretation and understanding, as well as acquisition of language.
12. This terminology seems to me misleading for reasons already discussed; we have little reason to believe that behavior is "caused" in

any well-understood sense of this term. I will, nevertheless, use the terminology, as is conventional, although with this proviso noted.

13. See Dummett (1981). For discussion of Quine's critique and others, see Chomsky (1975b). Dummett's comments appear in a review of Chomsky (1980b); the latter contains a critique of the proposal he outlines and of his own views on the theory of meaning.

14. See Searle's contribution and my response in the peer review journal *Behavioral and Brain Sciences*, 3, 1-61 (1980); see also the further discussion there by various commentators on excerpts from Chomsky (1980b) and my response; see also Chomsky (1980b). See also Searle's unpublished commentary on a version of these remarks presented at the Sloan Conference on Philosophy and Psychology, MIT, May 1984. I am indebted to Searle for clarification of his views in personal communication.

15. We might seek to determine more closely just what kind of "entering into" is relevant here, but although possibly an interesting question, it is not germane in this context.

16. Similar questions arise with no possibility of asking the agent; see Chomsky (1980b), pp. 102-103.

17. See Chomsky (1977, p. 207), and (1981, p. 2).

18. There might be a reasonable interpretation of the latter quest, not strictly relevant to the discussion here; see note 15.

19. On some confusions about the nature and force of the evidence, and the belief that only some categories of evidence serve to confer a mysterious property called "psychological reality," see Chomsky (1980b). See also Gilbert Harman's comments in the reference of note 14.

20. See references of note 14 and Chomsky (1975b).

21. Pylyshyn argues in contrast that we can speak of rules and representations only at the semantic-intentional level. The conclusion seems to me unsound, in fact hardly more than a dubious terminological proposal, although the analysis that lies behind it is informative and enlightening.

22. See Miller and Chomsky (1963), for some early speculation along these lines, and Berwick and Weinberg (1984) for more recent discussion. There is little reason to take for granted that properties of the language faculty have been specifically selected; see Chomsky (1965, 1968). Certainly evolutionary biology is not committed to such a view in general.

23. See McGinn (1981).

24. See White (1982) for a thoughtful and informative analysis of the commonsense concept, although I would question his conclusions about the relation of knowledge to ability for reasons discussed earlier; see pp. 9-13.

25. In particular, the book he was reviewing, Chomsky (1980b).

26. On the first, see Nagel (1969).

27. See, among others, Gruber (1976), Fillmore (1968), Jackendoff (1972, 1984), J.M. Anderson (1972), and Moravcsik (1975). On the possible relevance of sign language in identifying some of these elements and distinguishing their roles, see Gee and Kegl (1982), and Kegl (1984).

28. On this matter see Chomsky (1975b, 1980b) and Fodor (1983).