2 Explaining language use

In his John Locke lectures, Hilary Putnam argues "that certain human abilities – language speaking is the paradigm example – may not be theoretically explicable in isolation," apart from a full model of "human functional organization," which "may well be *unintelligible* to humans when stated in any detail." The problem is that "we are not, realistically, going to get a detailed explanatory model for the natural kind 'human being'," not because of "*mere* complexity" but because "we are partially opaque to ourselves, in the sense of *not* having the ability to understand one another as we understand hydrogen atoms." This is a "*constitutive* fact" about "human beings in the present period," though perhaps not in a few hundred years (Putnam 1978).

The "natural kinds" *human being* and *hydrogen atom* thus call for different kinds of inquiry, one leading to "detailed explanatory models," the other not, at least for now. The first category is scientific inquiry, in which we seek intelligible explanatory theories and look forward to eventual integration with the core natural sciences; call this mode of inquiry "naturalistic," focusing on the character of work and reasonable goals, in abstraction from actual achievement. Beyond its scope, there are issues of the scale of full "human functional organization," not a serious topic for (current) naturalistic inquiry but more like the study of everything, like attempts to answer such pseudo-questions as "how do things work?" or "why do they happen?" Many questions – including those of greatest human significance, one might argue – do not fall within naturalistic inquiry; we approach them in other ways. As Putnam stresses, the distinctions are not sharp, but they are useful nonetheless.

In a critical discussion of "sophisticated mentalism of the MIT variety" (specifically, Jerry Fodor's "language of thought"; Fodor 1975), Putnam adds some complementary observations on theoretical inquiry that would *not* help to explain language speaking. He considers the possibility that the brain sciences might discover that when we "think the word *cat*" (or a Thai speaker thinks the equivalent), a configuration C is formed in the brain. "This is fascinating if true," he concludes, perhaps a significant contribution to psychology and the brain sciences, "but what is its relevance to a discussion of the *meaning* of *cat*" (or of the Thai equivalent, or of C)? – the implication being that there is no relevance (Putnam 1988a).

We thus have two related theses. First, "language speaking" and other human abilities do not currently fall within naturalistic inquiry. Second, nothing could be learned about meaning (hence about a fundamental aspect of language speaking) from the study of configurations and processes of the brain (at least of the kind illustrated). The first conclusion seems to me understated and not quite properly formulated; the second, too strong. Let's consider them in turn.

The concept *human being* is part of our common-sense understanding, with properties of individuation, psychic persistence, and so on, reflecting particular human concerns, attitudes, and perspectives. The same is true of *language speaking*. Apart from improbable accident, such concepts will not fall within explanatory theories of the naturalistic variety; not just now, but ever. This is not because of cultural or even intrinsically human limitations (though these surely exist), but because of their nature. We may have a good deal to say about people, so conceived; even low-level accounts that provide weak explanation. But such accounts cannot be integrated into the natural sciences alongside of explanatory models for hydrogen atoms, cells, or other entities that we posit in seeking a coherent and intelligible explanatory model of the naturalistic variety. There is no reason to suppose that there is a "natural kind 'human being' "; at least if natural kinds are the kinds of nature, the categories discovered in naturalistic inquiry.

The question is not whether the concepts of common-sense understanding can themselves be studied in some branch of naturalistic inquiry; perhaps they can. Rather, it is whether in studying the natural world (for that matter, in studying these concepts, as part of the natural world), we view it from the standpoint provided by such concepts. Surely not. There may be scientific studies of some aspects of what people are and do, but they will not use the common-sense notions *human being* or *language speaking* – with their special role in human life and thought – in formulating their explanatory principles.

The same is true of common-sense concepts generally. Such notions as *desk* or *book* or *house*, let alone more "abstract" ones, are not appropriate for naturalistic inquiry. Whether something is properly described as a desk, rather than a table or a hard bed, depends on its designer's intentions and the ways we and others (intend to) use it, among other factors. Books are concrete objects. We can refer to them as such ("the book weighs five pounds"), or from an abstract perspective ("who wrote the book?"; "he wrote the book in his head, but then forgot about it"); or from both perspectives simultaneously ("the book he wrote weighed five pounds," "the book he is writing will weigh at least five pounds if it is ever published"). If I say "that deck of cards, which is missing a Queen, is too worn to use," that deck of cards is simultaneously taken to be a defective set and a strange sort of scattered "concrete object," surely not a mereological sum. The term *house* is used to refer to concrete objects, but from the standpoint of special human interests and goals and with curious properties. A house can be destroyed and rebuilt, like a city; London could be completely destroyed and rebuilt up the Thames in 1,000 years and still be London, under some circumstances. It is hard to imagine how these could be fit concepts for theoretical study of things, events, and processes in the natural world. Uncontroversially, the same is true of *matter*, *motion*, *energy*, *work*, *liquid*, and other common-sense notions that are abandoned as naturalistic inquiry proceeds; a physicist asking whether a pile of sand is a solid, liquid, or gas – or some other kind of substance – spends no time asking how the terms are used in ordinary discourse, and would not expect the answer to the latter question to have anything to do with natural kinds, if these are the kinds in nature (Jaeger and Nagel 1992).

It is only reasonable to expect that the same will be true of *belief*, *desire*, *meaning*, and *sound* of *words*, *intent*, etc., insofar as aspects of human thought and action can be addressed within naturalistic inquiry. To be an Intentional Realist, it would seem, is about as reasonable as being a Desk- or Sound-of-Language- or Cat- or Matter-Realist; not that there are no such things as desks, etc., but that in the domain where questions of realism arise in a serious way, in the context of the search for laws of nature, objects are not conceived from the peculiar perspectives provided by concepts of common-sense. It is widely held that "mentalistic talk and mental entities should eventually lose their place in our attempts to describe and explain the world" (Burge 1992). True enough, but it is hard to see the significance of the doctrine, since the same holds true, uncontroversially, for "physicalistic talk and physical entities" (to whatever extent the "mental"–"physical" distinction is intelligible).

Even the most elementary notions, such as *nameable thing*, crucially involve such intricate notions as human agency. What we take as objects, how we refer to them and describe them, and the array of properties with which we invest them, depend on their place in a matrix of human actions, interests, and intent in respects that lie far outside the potential range of naturalistic inquiry. The terms of language may also indicate positions in belief systems, which enrich further the perspectives these terms afford for viewing the world, though in ways inappropriate to the ends of naturalistic inquiry. Some terms – particularly those lacking "internal relational structure" (notably, so-called "natural kind terms") – may do little more than that, as far as the natural-language lexicon is concerned. (See, among others, Moravcsik 1975; Chomsky 1975b; Moravcsik 1990; Bromberger 1992a.) By "internal relational structure" I mean the selectional properties of such words as "give" (which takes an agent subject, theme object, and goal indirect object), lacking for "cat," "liquid," etc. The concepts of natural language, and common-sense generally, are not even candidates for naturalistic theories.

Putnam extends his conclusions to Brentano's thesis that "intentionality won't be reduced and won't go away": "there is no scientifically describable property that all cases of any particular intentional phenomenon have in common" (say, thinking about cats) (Putnam 1988a). More generally, intentional phenomena relate to people and what they do as viewed from the standpoint of human interests and unreflective thought, and thus will not (so viewed) fall within naturalistic theory, which seeks to set such factors aside. Like falling bodies, or the heavens, or liquids, a "particular intentional phenomenon" may be associated with some amorphous region in a highly intricate and shifting space of human interests and concerns. But these are not appropriate concepts for naturalistic inquiry.

We may speculate that certain components of the mind (call them the "science-forming faculty," to dignify ignorance with a title) enter into naturalistic inquiry, much as the language faculty (about which we know a fair amount) enters into the acquisition and use of language. The products of the science-forming faculty are fragments of theoretical understanding, naturalistic theories of varying degrees of power and plausibility involving concepts constructed and assigned meaning in a considered and determinate fashion, as far as possible, with the intent of sharpening or otherwise modifying them as more comes to be understood. Other faculties of the mind yield the concepts of common-sense understanding, which enter into natural-language semantics and belief systems. These simply "grow in the mind," much in the way that the embryo grows into a person. How sharp the distinctions may be is an open question, but they appear to be real nevertheless.

Sometimes there is a resemblance between concepts that arise in these different ways; possibly naturalistic inquiry might construct some counterpart to the common-sense notion *human being*, as H_20 has a rough correspondence to *water* (though earth, air, and fire, on a par with water for the ancients, lack such counterparts). It is a commonplace that any similarities to common-sense notions are of no consequence for science. It is, for example, no requirement for biochemistry to determine at what point in the transition from simple gases to bacteria we find the "essence of life"; and if some such categorization were imposed, the correspondence to some common-sense notion would matter no more than for (topological) *neighborhood*, *energy*, or *fish*.

Similarly, it is no concern of the psychology-biology of organisms to deal with such technical notions of philosophical discourse as *perceptual content*, with its stipulated properties (sometimes dubiously attributed to "folk psychology," a construct that appears to derive in part from parochial cultural conventions and traditions of academic discourse). Nor must these inquiries assign a special status to veridical perception under "normal" conditions. Thus, in the study of determination of structure from motion, it is immaterial whether the external event is successive arrays of flashes on a tachistoscope that yield the visual experience of a cube rotating in space, or an actual rotating cube, or stimulation of the retina, or optic nerve, or visual cortex. In any case, "the computational investigation concerns the nature of the internal representations used by the visual system and the processes by which they are derived" (Ullman 1979: 3), as does the study of algorithms and mechanisms in this and other work along lines pioneered by David Marr (1982). It is also immaterial whether people might accept the nonveridical cases as "seeing a cube" (taking "seeing" to be having an experience, whether "as if" or veridical); or whether concerns of philosophical theories of intentional attribution are addressed. A "psychology" dealing with the latter concerns would doubtless not be individualistic, as Martin Davies (1991) argues, but it would also depart from naturalistic inquiry into the nature of organisms, and possibly from authentic folk psychology as well.[1] To take another standard example, on the (rather implausible) assumption that a naturalistic approach to, say, jealousy were feasible, it is hardly likely that it would distinguish between states involving real or imagined objects. If "cognitive science" is taken to be concerned with intentional attribution, it may turn out to be an interesting pursuit (as literature is), but it is not likely to provide explanatory theory or to be integrated into the natural sciences.

As understanding progresses and concepts are sharpened, the course of naturalistic inquiry tends towards theories in which terms are divested of distorting residues of common-sense understanding, and are assigned a relation to posited entities and a place in a matrix of principles: *real number*, *electron*, and so on. The divergence from natural language is two-fold: the constructed terms abstract from the intricate properties of natural-language expressions; they are assigned semantic properties that may well not hold for natural language, such as reference (we must beware of what Strawson once called "the myth of the logically proper name," in natural language, and related myths concerning indexicals and pronouns; P. Strawson 1952: 216). As this course is pursued, the divergence from natural language increases; and with it, the divergence between the ways we understand *hydrogen atom*, on the one hand, and *human being* (*desk*, *liquid*, *heavens*, *fall*, *chase*, *London*, *this*, etc.), on the other.

But even a strengthened version of Putnam's first thesis does not entitle us to move on to the second, more generally, to conclude that naturalistic theories of the brain are of no relevance to understanding what people do. Under certain conditions, people see tachistoscopic presentations as a rotating cube or light moving in a straight line. A study of the visual cortex might provide understanding of why this happens, or why perception proceeds as it does in ordinary circumstances. And comparable inquiries might have a good deal to say about "language speaking" and other human activities.

Take Putnam's case: the discovery that thinking of cats evokes C. Surely such a discovery might have some relevance to inquiry into what Peter means (or refers to, or thinks about) when he uses the term *cat*, hence to "a discussion of the meaning of *cat*." For example, there has been a debate – in which Putnam has taken part – about the referential properties of *cat* if cats were found to be robots controlled from Mars. Suppose that after Peter comes to believe this, his brain does, or does not, form C when he refers to cats (thinks about them, etc.). That might be relevant to the debate. Or, take a realistic case: recent studies of electrical activity of the brain (event-related potentials, ERPs) show distinctive responses to nondeviant and deviant expressions and, among the latter, to violations of:

1. word meaning expectancies;
2. phrase-structure rules;
3. the specificity-of-reference condition on extraction of operators; and
4. locality conditions on movement (Neville *et al.* 1991).

Such results surely might be relevant to the study of the use of language, in particular, the study of meaning.

We can proceed further. Patterns of electrical activity of the brain correlate with the five categories of structure noted: nondeviance, and four types of deviance. But the study of these categories is also a study of the brain, its states and properties, just as study of algorithms involved in seeing a straight line or in doing long division is a study of the brain. Like other complex systems, the brain can be studied at various levels: atoms, cells, cell assemblies, neural networks, computational–representational (C–R) systems, etc. The ERP study relates two such levels: electrical activity of the brain and C–R systems. The study of each level is naturalistic both in the character of the work and in that integration with the core natural sciences is a prospect that can be reasonably entertained. In the context of Putnam's discussion, discoveries about the brain at these levels of inquiry are on a par with a discovery about the (imagined) configuration C, when Peter thinks of cats.

In the case of language, the C–R theories have much stronger empirical support than anything available at other levels, and are far superior in explanatory power; they fall within the natural sciences to an extent that inquiry into "language speaking" at the other levels does not. In fact, the current significance of the ERP studies lies primarily in their correlations with the much richer and better-grounded C–R theories. Within the latter, the five categories have a place and, accordingly, a wide range of indirect empirical support; in isolation from C–R theories, the ERP observations are just curiosities, lacking a theoretical matrix. Similarly, the discovery that C correlates with use of *cat* would, as an isolated fact, be more of a discovery about C than about the meaning of *cat* – and for that reason alone would shed little light on the controversy about robots controlled from Mars. To take another case, the discovery of perceptual displacement of clicks to phrase boundaries is, for now, more of a discovery about the validity of the experiment than about phrase boundaries. The reason is that evidence of other sorts about phrase boundaries – sometimes called "linguistic" rather than "psychological" evidence (a highly misleading terminology) – is considerably more compelling and embedded in a much richer explanatory structure. If click experiments were found to be sufficiently reliable in identifying the entities postulated in C–R theories, and if their theoretical framework were deepened, one might rely on them in cases where "linguistic evidence" is indecisive; possibly more, as inquiry progresses. (On some misunderstandings of these matters see Chapter 3 of this volume; Chomsky 1991a; 1991b).

For the present, the best-grounded naturalistic theories of language and its use are C–R theories. We assume, essentially on faith, that there is some kind of description in terms of atoms and molecules, though without expecting operative principles and structures of language and thought to be discernible at these levels. With a larger leap of faith, we tend to assume that there is an account in neurological terms (rather than, say, glial or vascular terms, though a look at the brain reveals glial cells and blood as well as neurons.[2] It may well be that the relevant elements and principles of brain structure have yet to be discovered. Perhaps C–R theories will provide guidelines for the search for such mechanisms, much as nineteenth-century chemistry provided crucial empirical conditions for radical revision of fundamental physics. The common slogan that "the mental is the neurophysiological at a higher level" – where C–R theories are placed within "the mental" – has matters backwards. It should be rephrased as the speculation that the neurophysiological may turn out to be "the mental at a lower level" – that is, the speculation that neurophysiology might, some day, prove to have some bearing on the "mental phenomena" dealt with in C–R theories. As for the further claims of eliminative materialism, the doctrine remains a mystery until some account is given of the nature of "the material"; and given that account, some reason why one should take it seriously or care if successful theories lie beyond its stipulated bounds.

For the present, C–R approaches provide the best-grounded and richest naturalistic account of basic aspects of language use. Within these theories, there is a fundamental concept that bears resemblance to the common-sense notion "language": the *generative procedure* that forms *structural descriptions* (SDs), each a complex of phonetic, semantic, and structural properties. Call this procedure an *I-language*, a term chosen to indicate that this conception of language is internal, individual, and intensional (so that distinct I-languages might, in principle, generate the same set of SDs, though the highly restrictive innate properties of the language faculty may well leave this possibility unrealized). We may take the linguistic expressions of a given I-language to be the SDs generated by it. A linguistic expression, then, is a complex of phonetic, semantic, and other properties. To have an I-language is something like having a "way to speak and understand," which is one traditional picture of what a language is. There is reason to believe that the I-languages ("grammatical competence") are distinct from conceptual organization and "pragmatic competence," and that these systems can be selectively impaired and developmentally dissociated (see Yamada 1990; John Marshall 1990).

The I-language specifies the form and meaning of such lexical elements as *desk*, *work*, and *fall*, insofar as these are determined by the language faculty itself. Similarly, it should account for properties of more complex expressions: for example, the fact that "John rudely departed" may mean either that he departed in a rude manner or that it was rude of him to depart, and that, in either case, he departed (perhaps an event semantics should be postulated as a level of representation to deal with such facts; see Higginbotham 1985; 1989). And it should explain the fact that the understood subject of *expect* in example (1) depends on whether *X* is null or is *Bill*, with a variety of other semantic consequences:

(1) John is too clever to expect anyone to talk to *X*.

And for the fact that, in my speech, *ladder* rhymes with *matter* but *madder* doesn't. In a wide range of such cases, nontrivial accounts are forthcoming. The study of C–R systems provides no little insight into how people articulate their thoughts and interpret what they hear, though of course it is as little – and as much – a study of these actions as the physiology and psychology of vision are studies of humans seeing objects.

A deeper inquiry into I-languages will seek to account for the fact that Peter has the I-language L_P while Juan has the I-language L_J – these statements being high-level abstractions, because in reality what Peter and Juan have in their heads is about as interesting for naturalistic inquiry as the course of a feather on a windy day. The basic explanation must lie in the properties of the language faculty of the brain. To a good approximation, the genetically-determined initial state of the language faculty is the same for Peter, Juan, and other humans. It permits only a restricted variety of I-languages to develop under the triggering and shaping effect of experience. In the light of current understanding, it is not implausible to speculate that the initial state determines the computational system of language uniquely, along with a highly structured range of lexical possibilities and some options among "grammatical elements" that lack substantive content. Beyond these possibilities, variation of I-languages may reduce to Saussurean arbitrariness (an association of concepts with abstract representations of sound) and parts of the sound system, relatively accessible and, hence, "learnable" (to use a term with misleading connotations). Small differences in an intricate system may, of course, yield large phenomenal differences, but a rational Martian scientist studying humans might not find the difference between English and Navajo very impressive.

The I-language is a (narrowly described) property of the brain, a relatively stable element of transitory states of the language faculty. Each linguistic expression (SD) generated by the I-language includes instructions for performance systems in which the I-language is embedded. It is only by virtue of its integration into such performance systems that this brain state qualifies as a language. Some other organism might, in principle, have the same I-language (brain state) as Peter, but embedded in performance systems that use it for locomotion. We are studying a real object, the language faculty of the brain, which has assumed the form of a full I-language and is integrated into performance systems that play a role in articulation, interpretation, expression of beliefs and desires, referring, telling stories, and so on. For such reasons, the topic is the study of human language.

The performance systems appear to fall into two general types: articulatory–perceptual, and conceptual–intentional.[3] If so, it is reasonable to suppose that a generated expression includes two *interface levels*, one providing information and instructions for the articulatory–perceptual systems, the other for the conceptual–intentional systems. One interface is generally assumed to be phonetic representation (Phonetic Form, PF). The nature of the other is more controversial; call it LF ("Logical Form").

The properties of these systems, or their existence, are matters of empirical fact. One should not be misled by unintended connotations of such terms as "logical form" and "representation," drawn from technical usage in different kinds of inquiry. Similarly, though there is a hint of the notions "deep grammar" and "surface grammar" of philosophical analysis, the concepts do not closely match. What is "surface" from the point of view of I-language is, if anything, PF, the interface with articulatory–perceptual systems. Everything else is "deep." The surface grammar of philosophical analysis has no particular status in the empirical study of language; it is something like phenomenal judgment, mediated by schooling, traditional authorities and conventions, cultural artifacts, and so on. Similar questions arise with regard to what is termed, much too casually, "folk psychology," as noted. One should regard such notions with caution: much may be concealed behind apparent phenomenal clarity.

The complex of I-language and performance systems enters into human action. It is an appropriate subject matter for naturalistic theories, which might carry us far towards understanding how and why people do what they do, though always falling short of a full account, just as a naturalistic theory of the body would fail to capture fully such human actions or achievements as seeing a tree or taking a walk.

Correspondingly, it would be misleading, or worse, to say that some part of the brain or an abstract model of it (for example, a neural net or programmed computer) sees a tree or figures out square roots. People in an ambiguous range of standard circumstances pronounce words, refer to cats, speak their thoughts, understand what others say, play chess, or whatever; their brains don't and computer programs don't – though study of brains, possibly with abstract modelling of some of their properties, might well provide insight into what people are doing in such cases. An algorithm constructed in a C–R theory might provide a correct account of what is happening in the brain when Peter sees a straight line or does long division or "understands Chinese,"[4] and might be fully integrated into a well-grounded theory at some other level of explanation (say, cells). But the algorithm, or a machine implementing it, would not be carrying out these actions, though we might decide to modify existing usage, as when we say that airplanes fly and submarines set sail (but do not swim). Nothing of substance is at stake. Similarly, while it may be that people carry out the action by virtue of the fact that their brains implement the algorithm, the same people would not be carrying out the action if they were mechanically implementing the instructions, in the manner of a machine (or of their brains). It may be that I see a straight line (do long division, understand English, etc.) by virtue of the fact that my brain implements a certain algorithm; but if I, the person, carry out the instructions mechanically, mapping some symbolic representation of the input to a representation of the output, neither I nor I-plus-algorithm-plus-external memory sees a straight line (etc.), again, for uninteresting reasons.[5]

It would also be a mistake, in considering the nature of performance systems, to move at once to a vacuous "study of everything." As a case in point, consider Donald Davidson's discussion of Peter as an "interpreter," trying to figure out what Tom has in mind when he speaks. Davidson observes that Peter may well use any information, background assumption, guesswork, or whatever, constructing a "passing theory" for the occasion. Consideration of an "interpreter" thus carries us to full models of human functional organization. Davidson concludes that there is no use for "the concept of a language" serving as a "portable interpreting machine set to grind out the meaning of an arbitrary utterance"; we are led to "abandon . . . not only the ordinary notion of a language, but we have erased the boundary between knowing a language and knowing our way around in the world generally." Since "there are no rules for arriving at passing theories," we "must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases" (Davidson 1986b: 446). "There is no such thing as a language," a recent study of Davidson's philosophy opens, with his approval (Davidson 1986b; Ramberg 1989).

The initial observation about "passing theories" is correct, but the conclusions do not follow. A reasonable response to the observation – if our goal is to understand what humans are and what they do – is to try to isolate coherent systems that are amenable to naturalistic inquiry and that interact to yield some aspects of the full complexity. If we follow this course, we are led to the conjecture that there is a generative procedure that "grinds out" linguistic expressions with their interface properties, and performance systems that access these instructions and are used for interpreting and expressing one's thoughts.

What about "the idea of a clearly defined shared structure which language-users acquire and then apply to cases"? Must we also postulate such "shared structures," in addition to I-language and performance systems? It is often argued that such notions as common "public language" or "public meanings" are required to explain the possibility of communication or of "a common treasure of thoughts," in Gottlob Frege's sense (Frege 1892/1965: 71). Thus, if Peter and Mary do not have a "shared language," with "shared meanings" and "shared reference," then how can Peter understand what Mary says? (Interestingly, no one draws the analogous conclusion about "public pronunciation.") One recent study holds that linguists can adopt an I-language perspective only "at the cost of denying that the basic function of natural languages is to mediate communication between its speakers," including the problem of "communication between *time slices of an idiolect*" (so-called "incremental learning"; Fodor and Lepore 1992).[6]

But these views are not well founded. Successful communication between Peter and Mary does not entail the existence of shared meanings or shared pronunciations in a public language (or a common treasure of thoughts or articulations of them), any more than physical resemblance between Peter and Mary entails the existence of a public form that they share. As for the idea that "the basic function of natural languages is to mediate communication," it is unclear what sense can be given to an absolute notion of "basic function" for any biological system; and if this problem can be overcome, we may ask why "communication" is the "basic function." Furthermore, the transition problem seems no more mysterious than the problem of how Peter can be the person he is, given the stages through which he has passed. Not only is the I-language perspective appropriate to the problems at hand, but it is not easy to imagine a coherent alternative.

It may be that when he listens to Mary speak, Peter proceeds by assuming that she is identical to him, modulo M, some array of modifications that he must work out. Sometimes the task is easy, sometimes hard, sometimes hopeless. To work out M, Peter will use any artifice available to him, though much of the process is doubtless automatic and unreflective.[7] Having settled on M, Peter will, similarly, use any artifice to construct a "passing theory" – even if M is null. Insofar as Peter succeeds in these tasks, he understands what Mary says as being what he means by his comparable expression. The only (virtually) "shared structure" among humans generally is the initial state of the language faculty. Beyond that we expect to find no more than approximations, as in the case of other natural objects that grow and develop.

Discussion of language and language use regularly introduces other kinds of shared structure: communities with their languages, common languages across a broader culture, etc. Such practices are standard in ordinary casual discourse as well. Thus, we say that Peter and Tom speak the same language, but Juan speaks a different one. Similarly, we say that Boston is near New York, but not near London, or that Peter and Tom look alike, but neither looks like John. Or, we might reject any of these assertions. There is no right or wrong choice in abstraction from interests that may vary in every imaginable way. There are also no natural categories, no idealizations. In these respects, speaking the same language is on a par with being-near or looking-like. A standard remark in an undergraduate linguistics course is Max Weinreich's quip that a language is a dialect with an army and a navy, but dialects are also nonlinguistic notions, which can be set up one way or another, depending on particular interests and concerns. Such factors as conquests, natural barriers (oceans, mountains), national TV, etc. may induce illusions on this matter, but no notion of "common language" has been formulated in any useful or coherent way, nor do the prospects seem hopeful. Any approach to the study of language or meaning that relies on such notions is highly suspect.

Suppose, for example, that "following a rule" is analyzed in terms of communities: Jones follows a rule if he conforms to the practice or norms of the community. If the "community" is homogeneous, reference to it contributes nothing (the notions *norm*, *practice*, *convention*, etc. raise further questions). If the "community" is heterogeneous – apart from the even greater unclarity of the notion of norms (practice, etc.) for this case – several problems arise. One is that the proposed analysis is descriptively inaccurate. Typically, we attribute rule-following in the case of notable *lack* of conformity to prescriptive practice or alleged norms. Thus we might say that Johnny, who is three, is following his own rule when he says *brang* instead of *brought*; or that his father Peter is following the "wrong rule" ("violating the rules") when he uses *disinterested* to mean *uninterested* (as most people do). But only a linguist would say that Johnny and Peter are observing Condition (B) of the Binding theory (Chomsky 1981a: 188), as does the "community" generally (in fact, the community of all language speakers, very likely). The more serious objection is that the notion of "community" or "common language" makes as much sense as the notion "nearby city" or "look alike," without further specification of interests, leaving the analysis vacuous.[8]

For familiar reasons, nothing in this suggests that there is any problem in informal usage, any more than in the ordinary use of such expressions as *Boston is near New York* or *John is almost home*. It is just that we do not expect such notions to enter into explanatory theoretical discourse. They may be appropriate for informal discussion of what people do, with tacit assumptions of the kind that underlie ordinary discourse in particular circumstances; or even for technical discourse, where the relevant qualifications are tacitly understood. They have no further place in naturalistic inquiry, or in any attempt to sharpen understanding.

Alleged social factors in language use often have a natural individualist–internalist interpretation. If Peter is improving his Italian or Gianni is learning his, they are (in quite different ways) becoming more like a wide range of people; both the modes of approximation and selection of models vary with our interests. We gain no insight into what they are doing by supposing that there is a fixed entity that they are approaching, even if some sense can be made of this mysterious notion. If Bert complains of arthritis in his ankle and thigh, and is told by a doctor that he is wrong about both, but in different ways, he may (or may not) choose to modify his usage to that of the doctor's. Apart from further detail, which may vary widely with changing contingencies and concerns, nothing seems missing from this account. Similarly, ordinary talk of whether a person has mastered a concept requires no notion of common language. To say that Bert has not mastered the concept *arthritis* or *flu* is simply to say that his usage is not exactly that of people we rely on to cure us – a normal situation. If my neighbor Bert tells me about his arthritis, my initial posit is that he is identical to me in this usage. I will introduce modifications to interpret him as circumstances require; reference to a presumed "public language" with an "actual content" for *arthritis* sheds no further light on what is happening between us, even if some clear sense can be given to the tacitly assumed notions. If I know nothing about elms and beeches beyond the fact that they are large deciduous trees, nothing beyond this information might be represented in my mental lexicon (possibly not even that, as noted earlier); the understood difference in referential properties may be a consequence of a condition holding of the lexicon generally: lack of indication of a semantic relation is taken to indicate that it does not hold.[9]

Questions remain – factual ones, I presume – as to just what kind of information is within the lexicon, as distinct from belief systems. Changes in usage, as in the preceding cases, may in fact be marginal changes of I-language, or changes in belief systems, here construed as (narrowly described) C–R systems of the mind, which enrich the perspectives and standpoints for thought, interpretation, language use and other actions (call them *I-belief systems*, some counterpart to beliefs that might be discovered in naturalistic inquiry). Work in lexical semantics provides a basis for empirical resolution in some cases (particularly in the verbal system, with its richer relational structure), keeping to the individualist–internalist framework.

Little is understood about the general architecture of the mind/brain outside of a few scattered areas, typically not those that have been the focus of the most general considerations of so-called "cognitive science." There has, for example, been much interesting discussion about a theory of belief and its possible place in accounting for thought and action. But substantive empirical work that might help in examining, refining, or testing these ideas is scarcely available. It seems reasonable at least to suppose that I-beliefs do not form a homogeneous set; the system has further structure that may provide materials for decisions about false belief and misidentification. Suppose that some I-beliefs are *identifying* beliefs and others not, or that they range along such a spectrum, where the latter (or the lesser) are more readily abandoned without affecting conditions for referring. Suppose, say, that Peter's information about Martin van Buren is exhausted by the belief that he was (1) the President of the United States and (2) the sixteenth President, (1) being more of an identifying belief than (2). If Peter learns that Lincoln was the sixteenth President he might drop the nonidentifying I-belief while using the term to refer. If he is credibly informed that all the history books are mistaken and van Buren wasn't a President at all, he is at a loss as to how to proceed. That seems a reasonable first step towards as much of an analysis as an internalist perspective can provide, and as much as seems factually at all clear. Further judgments can sometimes be made in particular circumstances, in varied and conflicting ways.[10]

It may be that a kind of public (or interpersonal) character to thought and meaning results from uniformity of initial endowment, which permits only I-languages that are alike in significant respects, thus providing some empirical reason to adopt some version of the Fregean doctrine that "it cannot well be denied that mankind possesses a common treasure of thoughts which is transmitted from generation to generation" (Frege 1892/1965: 71). And the special constructions of the science-forming faculty may also approach a public character (more to the point, for Frege's particular concerns). But for the systems that grow naturally in the mind, beyond the instantiation of initial endowment as I-language (perhaps also I-belief and related systems), the character of thought and meaning varies as interest and circumstance vary, with no clear way to establish further categories, even ideally. Appeals to a common origin of language or speculations about natural selection, which are found throughout the literature, seem completely beside the point.

Consider the shared initial state of the language faculty of the brain, and the limited range of I-languages that are attainable as it develops in early life. When we inquire into lexical properties, we find a rich texture of purely internalist semantics, with interesting general properties, and evidence for formal semantic relations (including analytic connections; see references on p. 22). Furthermore, a large part of this semantic structure appears to derive from our inner nature, determined by the initial state of our language faculty, hence unlearned and universal for I-languages. Much the same is true of phonetic and other properties. In short, I-language (including internalist semantics) seems much like other parts of the biological world.

We might well term all of this a form of syntax, that is, the study of the symbolic systems of C–R theories ("mental representation"). The same terminology remains appropriate if the theoretical apparatus is elaborated to include mental models, discourse representations, semantic values, possible worlds as commonly construed, and other theoretical constructions that still must be related in some manner to things in the world; or to the entities postulated by our science-forming faculty, or constructed by other faculties of the mind.

The internally-determined properties of linguistic expressions can be quite far-reaching, even in very simple cases. Consider again the word *house*, say, in the expression *John is painting the house brown*, a certain collection of structural, phonetic, and semantic properties. We say it is the same expression for Peter and Tom only in the sense in which we might say that their circulatory or visual systems are the same: they are similar enough for the purposes at hand. One structural property of the expression is that it consists of six words. Other structural properties differentiate it from *John is painting the brown house*, which has correspondingly different conditions of use. A phonetic property is that the last two words, *house* and *brown*, share the same vowel; they are in the formal relation of assonance, while *house* and *mouse* are in the formal relation of rhyme, two relations on linguistic expressions definable in terms of their phonological features.[11] A semantic property is that one of the two final words can be used to refer to certain kinds of things, and the other expresses a property of these. Here, too, there are formal relations expressible in terms of features of the items, for example, between *house* and *building*. Or, to take a more interesting property, if John is painting the house brown, then he is applying paint to its exterior surface, not its interior; a relation of entailment holds between the corresponding linguistic expressions.

Viewed formally, relations of entailment have much the same status as rhyme; they are formal relations among expressions, which can be characterized in terms of their linguistic features. Certain relations happen to be interesting ones, as distinct from many that are not, because of the ways I-languages are embedded in performance systems that use these instructions for various human activities.

Some properties of the expression are universal, others language-particular. It is a universal phonetic property that the vowel of *house* is shorter than the vowel of *brown*; it is a particular property that the vowel in my I-language is front rather than mid, as it is in some I-languages similar to mine. The fact that a brown house has a brown exterior, not interior, appears to be a language universal, holding of "container" words of a broad category, including ones we might invent: *box*, *airplane*, *igloo*, *lean-to*, etc. To paint a spherical cube brown is to give it a brown exterior. The fact that *house* is distinguished from *home* is a particular feature of the I-language. In English, I return to my home after work; in Hebrew, I return to the house.

When we move beyond lexical structure, conclusions about the richness of the initial state of the language faculty, and its apparently special structure, are reinforced. Consider such expressions as those in example (2):

(2) a He thinks the young man is a genius.
b The young man thinks he is a genius.
c His mother thinks the young man is a genius.

In (2b) or (2c), the pronoun may be referentially dependent on *the young man*; in (2a) it cannot (though it might be used to refer to the young man in question, an irrelevant matter). The principles underlying these facts appear to be universal, at least in large measure;[12] again, they yield rich conditions on semantic interpretation, on intrinsic relations of meaning among expressions, including analytic connections. Furthermore, in this domain we have theoretical results of some depth, with surprising consequences. Thus, the same principles appear to yield the semantic properties of expressions of the form of example (1), on page 27.

Given the performance systems, the representation at the interface level PF imposes restrictive conditions on use (articulation and perception, in this case). The same is true of the LF representation, as illustrated in examples (1) and (2), or at the lexical level, in the special status of the exterior surface for container words. A closer look reveals further complexity. The exterior surface is distinguished in other ways within I-language semantics. If I see the house, I see its exterior surface; seeing the interior surface does not suffice. If I am inside an airplane, I see it only if I look out the window and see the surface of the wing, or if there is a mirror outside that reflects its exterior surface. But the house is not just its exterior surface, a geometrical entity. If Peter and Mary are equidistant from the surface – Peter inside and Mary outside – Peter is not near the house, but Mary might be, depending on the current conditions for nearness. The house can have chairs inside it or outside it, consistent with its being regarded as a surface. But while those outside may be near it, those inside are necessarily not. So the house involves its exterior surface and its interior. But the interior is abstractly conceived; it is the same house if I fill it with cheese or move the walls – though if I clean the house I may interact only with things in the interior space, and I am referring only to these when I say that the house is a mess or needs to be redecorated. The house is conceived as an exterior surface and an interior space (with complex properties). Of course, the house itself is a concrete object; it can be made of bricks or wood, and a wooden house does not just have a wooden exterior. A brown wooden house has a brown exterior (adopting the abstract perspective) and is made of wood (adopting the concrete perspective). If my house used to be in Philadelphia, but is now in Boston, then a physical object was moved. In contrast, if my home used to be in Philadelphia, but is now in Boston, then no physical object need have moved, though my home is also concrete – though in some manner also abstract, whether understood as the house in which I live, or the town, or country, or universe; a house is concrete in a very different sense. The *house – home* difference has numerous consequences: I can go home, but not go house; I can live in a brown house, but not a brown home; in many languages, the counterpart of *home* is adverbial, as partially in English too.

Even in this trivial example, we see that the internal conditions on meaning are rich, complex, and unsuspected; in fact, barely known. The most elaborate dictionaries do not dream of such subtleties; they provide no more than hints that enable the intended concept to be identified by those who already have it (at least, in essential respects). The I-variant of Frege's telescope operates in curious and intricate ways.

There seems at first glance to be something paradoxical in these descriptions. Thus, houses and homes are concrete but, from another point of view, are considered quite abstractly, though abstractly in very different ways; similarly, books, decks of cards, cities, etc. It is not that we have confused ideas – or inconsistent beliefs – about houses and homes, or boxes, airplanes, igloos, spherical cubes, etc. Rather, a lexical item provides us with a certain range of perspectives for viewing what we take to be the things in the world, or what we conceive in other ways; these items are like filters or lenses, providing ways of looking at things and thinking about the products of our minds. The terms themselves do not refer, at least if the term *refer* is used in its natural-language sense; but people can use them to refer to things, viewing them from particular points of view – which are remote from the standpoint of the natural sciences, as noted.

The same is true wherever we inquire into I-language. London is not a fiction, but considering it as London – that is, through the perspective of a city name, a particular type of linguistic expression – we accord it curious properties: as noted earlier, we allow that under some circumstances, it could be completely destroyed and rebuilt somewhere else, years or even millennia later, still being London, that same city. Charles Dickens described Washington as "the City of Magnificent Intentions," with "spacious avenues, that begin in nothing, and lead nowhere; streets, mile-long, that only want houses, roads, and inhabitants; public buildings that need but a public to be complete; and ornaments of great thoroughfares, which only lack great thoroughfares to ornament" – but still Washington. We can regard London with or without regard to its population: from one point of view, it is the same city if its people desert it; from another, we can say that London came to have a harsher feel to it through the Thatcher years, a comment on how people act and live. Referring to London, we can be talking about a location or area, people who sometimes live there, the air above it (but not too high), buildings, institutions, etc., in various combinations (as in *London is so unhappy, ugly, and polluted that it should be destroyed and rebuilt 100 miles away*, still being the same city). Such terms as *London* are used to talk about the actual world, but there neither are nor are believed to be things-in-the-world with the properties of the intricate modes of reference that a city name encapsulates. Two such collections of perspectives can fit differently into Peter's system of beliefs, as in Kripke's puzzle. (For extensive discussion from a somewhat similar point of view, see Bilgrami 1992.)

For purposes of naturalistic inquiry, we construct a picture of the world that is dissociated from these "common-sense" perspectives (never completely, of course; we cannot become something other than the creatures we are[13]). If we intermingle such different ways of thinking about the world, we may find ourselves attributing to people strange and even contradictory beliefs about objects that are to be regarded somehow apart from the means provided by the I-language and the I-belief systems that add further texture to interpretation. The situation will seem even more puzzling if we entertain the obscure idea that certain terms have a relation to things ("reference") fixed in a common public language, which perhaps even exists "independently of any particular speakers," who have a "partial, and partially erroneous, grasp of the language" (Dummett 1986); and that these "public-language terms" in the common language refer (in some sense to be explained) to such objects as London taken as a thing divorced from the properties provided by the city name (or some other mode of designation) in a particular I-language, and from the other factors that enter into Peter's referring to London. Problems will seem to deepen further if we abstract from the background of individual or shared beliefs that underlie normal language use. All such moves go beyond the bounds of a naturalistic approach, some of them, perhaps, beyond sensible discourse.

They also go beyond internalist limits, which is a different matter. A naturalistic approach does not impose internalist, individualist limits. Thus, if we study (some counterpart to) persons as phases in the history of ideally immortal germ cells, or as stages in the conversion of oxygen to carbon dioxide, we depart from such limits. But if we are interested in accounting for what people do, and why, insofar as that is possible through naturalistic inquiry, the argument for keeping to these limits seems persuasive.[14]

We began by considering the (hypothetical) discovery that Peter's brain produces the configuration C when he thinks about cats. We then moved to the more realistic example of ERPs, and the still more realistic case (from a scientific standpoint) of C–R systems; one may think of their elements as on a par with C, though now real, not hypothetical, we have reason to believe. The same would be true of a naturalistic approach that departs from these internalist limits, viewing Peter's brain as part of a larger system of interactions. The analogy would no longer be to the configuration C produced in Peter's brain when he thinks of cats, but to some physical configuration C′ involving C along with something else, perhaps something about cats. We are now in the domain of the hypothetical – I know of no serious candidate. But suppose that such an approach can be devised and proves to yield insight into questions of language use. If so, that might modify the ways we study language and psychology, but would not bridge the gap to an account of people and what they do.

We have to distinguish between a hypothetical externalist naturalism of the kind just sketched, and nonnaturalist externalism that attempts to treat human action (referring to or thinking about cats, etc.) in the context of communities, real or imagined things in the world, and so on. Such approaches are to be judged on their merits, as efforts to make some sense out of questions that lie beyond naturalistic inquiry – like questions about energy, falling stones, the heavens, etc. – in the ordinary sense of the terms. I have mentioned some reason for skepticism about recourse to communities and their practices, or public languages with public meanings. Consider further the other facet of externalism, an alleged relation between words and things.

Within internalist semantics, there are explanatory theories of considerable interest that are developed in terms of a relation *R* (read "refer") that is postulated to hold between linguistic expressions and something else, entities drawn from some stipulated domain D (perhaps semantic values).[15]

The relation *R*, for example, holds between the expressions *London* (*house*, etc.) and entities of D that are assumed to have some relation to what people refer to when they use the words *London* (*house*, etc.), though that presumed relation remains obscure. As noted, I think such theories should be regarded as a variety of syntax. The elements they postulate are on a par, in the respects relevant here, with phonological or phrase-structure representations, or the hypothetical brain configuration C; we might well include D and *R* within the SD (the linguistic expression), as part of an interface level.

Explanation of the phenomena of example (2) (on page 35) is commonly expressed in terms of the relation *R*. The same theories of binding and anaphora carry over without essential change if we replace *young* in example (2) by *average*, *typical*, or replace *the young man* by *John Doe*, stipulated to be the average man for the purposes of a particular discourse.[16] The same theories also carry over to anaphoric properties of the pronouns in examples (3) and (4):

(3) a It brings good health's rewards.
b Good health brings its rewards.
c Its rewards are what make good health worth striving for.

(4) a [There is a flaw in the argument], but it was quickly found.
b [The argument is flawed], but it was quickly found.

In terms of the relation *R*, stipulated to hold between *the average man*, *John Doe*, *good health*, *flaw*, and entities drawn from D, we can account for the differential behavior of the pronoun exactly as we would with *the young man*, *Peter*, *fly* ("there is a fly in the coffee"). The relations of anaphora differ in (4a and 4b), though there is no relevant difference in meaning between the bracketed clauses. And it might well turn out that these expressions, along with such others as "the argument has a flaw" (with the anaphoric options of (4a)), share still deeper structural properties, possibly even the same structural representation at the level relevant to the internal semantics of the phrases, a possibility that has been explored for some years (see Tremblay 1991).[17] The same is true in more exotic cases. It would seem perverse to seek a relation between entities in D and things in the world – real, imagined, or whatever – at least, one of any generality. One may imagine that the relation of elements of D to things in the world is more "transparent" than in the case of other syntactic representations, as the relation to sound waves is more "transparent" for phonetic than for phonological representation; but even if so, these studies do not pass beyond the syntax of mental representations. The relation *R* and the construct D must be justified on the same kinds of grounds that justify other technical syntactic notions; that is, those of phonology, or the typology of empty categories in syntax. An occasional resemblance between *R* and the term *refer* of ordinary language has no more significance than it would in the case of *momentum* or *undecidability*.

Specifically, we have no intuitions about *R*, any more than we do about *momentum* or *undecidability* in the technical sense, or about *c-command* or *autosegmental* in (other parts) of the C–R theories of syntax[18]; the terms have the meanings assigned to them. We have intuitive judgments about the notion used in such expressions as *Mary often refers to the young man as a friend (to the average man as John Doe, to good health as life's highest goal)*. But we have no such intuitions about the relation *R* holding between *Mary* (or *the average man*, *John Doe*, *good health*, *flaw*) and postulated elements of D. *R* and D are what we specify that they are, within a framework of theoretical explanation. We might compare *R* and D to *P* and PF, where *P* is a relation holding between an expression and its PF representation (between "took" and [t^huk], perhaps), though in the latter case the concepts fit into a much better-grounded and richer theory of interface relations.

Suppose that postulation of *R* and D is justified by explanatory success within the C–R theory of I-language, alongside of *P* and PF, *c-command*, and *autosegmental*. That result lends no support to the belief that some *R*-like relation, call it *R′*, holds between words and things, or things as they are imagined to be, or otherwise conceived. Postulation of such a relation would have to be justified on some grounds, as in the case of any other invented technical notion. And if we devise a relation *R′* holding between linguistic expressions and "things," somehow construed, we would have no intuitions about it – matters become only more obscure if we invoke unexplained notions of "community" or "public language," taken in some absolute sense. We do have intuitive judgments concerning linguistic expressions and the particular perspectives and points of view they provide for interpretation and thought. Furthermore, we might proceed to study how these expressions and perspectives enter into various human actions, such as referring. Beyond that, we enter the realm of technical discourse, deprived of intuitive judgment.

Take Putnam's influential Twin-Earth thought experiment (Putnam 1975). We can have no intuitions as to whether the term *water* has the same "reference" for Oscar and twin-Oscar: that is a matter of decision about the new technical term "reference" (some particular choice for *R′*). We have judgments about what Oscar and twin-Oscar might be referring to, judgments that seem to vary considerably as circumstances vary. Under some circumstances, Putnam's proposals about "same liquid," a (perhaps unknown) notion of the natural sciences, seem very plausible; under other circumstances, notions of sameness or similarity drawn from common-sense understanding seem more appropriate, yielding different judgments. It does not seem to me at all clear that there is anything general to say about these matters, or that any general or useful sense can be given to such technical notions as "wide content" (or any other notion fixing "reference") in any of the externalist interpretations.

If so, questions arise about the status of what Putnam, in his Locke lectures (Putnam 1988a: Chapter 2), calls the "social co-operation plus contribution of the environment theory of the *specification* of reference," a fuller and more adequate version of the "causal theory of reference" developed in his paper "The Meaning of 'Meaning'" (Putnam 1975) and Saul Kripke's *Naming and Necessity* (Kripke 1972), both now landmarks in the field.

"Social co-operation" has to do with "the division of linguistic labor": the role of experts in determining the reference of my terms *elm* and *beech*, for example. Putnam provides a convincing account for certain circumstances. Under some conditions, I would, indeed, agree that what I am referring to when I use the term *elm* is what is meant by an expert, perhaps an Italian gardener with whom I share only the Latin terms (though there is no meaningful sense in which we are part of the same "linguistic community" or speak a "common language"); under other conditions, probably not, but that is to be expected in an inquiry reaching as far as all of "human functional organization," virtually a study of everything. As mentioned earlier, it is not clear whether the question relates to I-language or I-belief, assuming the theoretical construction to be valid.

As for the "environment theory," it could contribute to specification of reference only if there were some coherent notion of "reference" (*R′*) holding between linguistic expressions and things, which is far from obvious, though people do use these expressions (in various ways) to refer to things, adopting the perspectives that these expressions provide. There are circumstances in which the particular conclusions usually drawn seem appropriate, in which "same species," "same liquid," etc., help determine what I am referring to; and there are other circumstances in which they do not.[19]

It also seems unclear that metaphysical issues arise in this context. To take some of Kripke's examples, doubtless there is an intuitive difference between the judgment that Nixon would be *the same person* if he had not been elected President of the USA in 1968, while he would not be the same person if he were not a person at all (say, if he were a silicon-based person replica). But that follows from the fact that *Nixon* is a personal name, offering a way of referring to Nixon *as a person*; it has no metaphysical significance. If we abstract from the perspective provided by natural language, which appears to have no pure names in the logician's sense (the same is true of variables, at least if pronouns are considered variables, and of indexicals, if we consider their actual conditions of use in referring), then intuitions collapse: Nixon would be a different *entity*, I suppose, if his hair were combed differently. Similarly, the object in front of me is not essentially a desk or a table; that very object could be any number of different things, as interests, functions, intentions of the inventor, etc. vary. To cite some recent work, Joseph Almog's judgment that the mountain Nanga Parbat is a mountain *essentially* might be intelligible under some circumstances; however, contrary to what he assumes, his "coherent–abstraction test" seems to me to permit us, under other circumstances, to deprive Nanga Parbat of this property, leaving it as the same entity: say, if the sea level rises high enough for its top to become an island, in which case it is no more a mountain than Britain is; or if earth is piled around it up to its peak, but a millimeter away, in which case it is not a mountain but part of a plateau surrounded by a crevice, though it remains the very same entity (Almog 1991).

In summary, it is questionable that standard conclusions can survive a closer analysis of the technical notions "reference" (in some *R*′-like sense) or "specification of reference." There may well be justification for the notion *R* internal to C–R theories (basically a syntactic notion, despite appearances). But there seems to be little reason to suppose that an analogous notion *R*′ can be given a coherent and useful formulation as a relation holding between expressions and some kind of things, divorced from particular conditions and circumstances of referring. If that is so, there will also be no reasonable inquiry into a notion of "sense" or "content" that "fixes reference" (*R*′), at least for natural language, though there is a promising (syntactic) inquiry into conditions for language use (including referring).

As discussed earlier, naturalistic inquiry may lead to the creation of language-like accretions to the I-language; for these, an *R*′-like notion may be appropriate, as terms are divested of the I-language properties that provide interpretive perspectives and semantic relations, are dissociated from I-belief, and are assigned properties lacking in natural language. These constructed systems may use resources of the I-language (pronunciation, morphology, sentence structure, etc.), or may transcend them (introducing mathematical formalisms, for example). The I-language is a product of the language faculty, abstracted from other components of the mind; this is an idealization of course, hence to be justified or rejected on the basis of its role in an explanatory framework. The picture could be extended, plausibly it seems, by distinguishing the system of common-sense belief from products of the science-forming faculty. The latter are neither I-languages nor I-belief systems, and for these it may well be appropriate to stipulate a relation *R′*.

Some of the motivation for externalist approaches derives from the concern to make sense of the history of science. Thus, Putnam argues that we should take the early Niels Bohr to have been referring to electrons in the quantum-theoretic sense, or we would have to "dismiss all of his 1900 beliefs as totally wrong," (Putnam 1988a) perhaps on a par with someone's beliefs about angels, a conclusion that is plainly absurd. The same is true of pre-Dalton chemists speaking of atoms. And perhaps, on the same grounds, we would say that chemists pre-Avogadro were referring to what we call atoms and molecules, though for them the terms were interchangeable, apparently.

The discussion assumes that such terms as *electron* belong to the same system as *house*, *water*, and pronominal anaphora, so that conclusions about *electron* carry over to notions in the latter category. That assumption seems to be implicit in Putnam's proposal that "To determine the intrinsic complexity of a task is to ask, *How hard is it in the hardest case*?," the "hardest case" for "same reference" or "same meaning" being posed by such concepts as *momentum* or *electron* in physics. But the assumption is dubious. The study of language should seek a more differentiated picture than that, and what is true of the technical constructions of the science-forming faculty might not hold for the natural-language lexicon. Suppose we grant the point nevertheless. Agreeing further that an interest in intelligibility in scientific discourse across time is a fair enough concern, still it cannot serve as the basis for a general theory of meaning; it is, after all, only one concern among many, and not a central one for the study of human psychology. Furthermore, there are internalist paraphrases. Thus we might say that in Bohr's earlier usage, he expressed beliefs that were literally false, because there was nothing of the sort he had in mind in referring to electrons; but his picture of the world and articulation of it was structurally similar enough to later conceptions so that we can distinguish his beliefs about electrons from beliefs about angels. What is more, that seems a reasonable way to proceed.

To take a far simpler example from the study of language, consider a debate some 30 years ago over the nature of phonological units. Structural phonologists postulated segments (phonemes) and phonetic features, with a certain collection of properties. Generative phonologists argued that no such entities exist, and that the actual elements have somewhat different properties. Suppose that one of these approaches looks correct (say, the latter). Were structural phonologists therefore referring all along to segments and features in the sense of generative phonology? Surely not. They flatly denied that, and were right to do so. Were they talking gibberish? Again, surely not. Structuralist phonology is intelligible; without any assumption that there are entities of the kind it postulated, much of the theory can be reinterpreted within generative phonology, with results essentially carried over. There is no principled way to determine how this is done, or to determine the "similarity of belief" between the two schools of thought or what thoughts and beliefs they shared. Sometimes it is useful to note resemblances and reformulate ideas, sometimes not. The same is true of the earlier and later Bohr. Nothing more definite is required to maintain the integrity of the scientific enterprise or a respectable notion of progress towards the truth about the world, insofar as it falls within human cognitive capacity.

It is worth noting that an analysis in these terms, eschewing externalist assumptions on fixation of reference, is consistent with the intuitions of respected figures. The discussion of the meaning of *electron*, *water*, etc. projects backwards in time, but we can project forward as well. Consider the question whether machines can think (understand, plan, solve problems, etc). By standard externalist arguments, the question should be settled by the truth about thought: what is the essence of Peter's thinking about his children, or solving a quadratic equation, or playing chess, or interpreting a sentence, or deciding whether to wear a raincoat? But that is not the way it seemed to Ludwig Wittgenstein and Alan Turing, to take two notable examples. For Wittgenstein, the question whether machines think cannot seriously be posed: "We can only say of a human being and what is like one that it thinks" (Wittgenstein 1958: 113), maybe dolls and spirits; that is the way the tool is used. Turing, in his classic 1950 paper, wrote that the question whether machines can think

> may be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (Turing 1950: 442)

Wittgenstein and Turing do not adopt the standard externalist account. For Wittgenstein, the questions are just silly: the tools are used as they are; and if the usage changes, the language has changed, the language being nothing more than the way we use the tools. Turing too speaks of the language of "general educated opinion" changing, as interests and concerns change. In our terms, there will be a shift from the I-languages that Wittgenstein describes to new ones, in which the old word *think* will be eliminated in favor of a new word that applies to machines as well as people. To ask in 1950 whether machines think is as meaningful as the question whether airplanes and people (say, high jumpers) really fly; in English airplanes do and high jumpers don't (except metaphorically), in Hebrew neither do, in Japanese both do. Such facts tell us nothing about the (meaningless) question posed, but only about marginal and rather arbitrary variations of I-language. The question of what *atom* meant pre-Dalton, or *electron* for Bohr in 1900, seems comparable, in relevant respects, to the question of what *think* meant for Wittgenstein and Turing; not entirely comparable, because *think*, *atom*, and *electron* should probably not be regarded as belonging to a homogeneous I-language. In all these cases, the internalist perspective seems adequate, not only to the intuitions of Wittgenstein and Turing, but to an account of what is transpiring; or what might happen as circumstances and interests vary.

Perhaps one might argue that recent semantic theories supersede the intuitions of Wittgenstein and Turing because of their explanatory success. That does not, however, seem a promising idea; explanatory success will hardly bear that burden. In general, we have little reason now to believe that more than a Wittgensteinian assembly of particulars lies beyond the domain of internalist inquiry, which is, however, far richer and informative than Wittgenstein, John Austin (1962), and others supposed.

Naturalistic inquiry will always fall short of intentionality. At least in these terms, "intentionality won't be reduced and won't go away," as Putnam puts it, and "language speaking" will remain not "theoretically explicable" (Putnam 1998a: 1). The study of C–R systems, including "internalist semantics," appears to be, for now, the most promising form of naturalistic inquiry, with a reasonably successful research program; understanding of performance systems is more rudimentary, but within the range of inquiry, in some respects at least. These approaches raise problems of the kind familiar throughout the natural sciences, but none that seem qualitatively different. Pursuing them, we can hope to learn a good deal about the devices that are used to articulate thoughts, interpret, and so on. They leave untouched many other questions, but it remains to be shown that these are real questions, not pseudo-questions that indicate topics of inquiry that one might hope to explore – but little more than that.