

Mind Association

Language and Nature

Author(s): Noam Chomsky

Source: *Mind*, New Series, Vol. 104, No. 413 (Jan., 1995), pp. 1-61

Published by: Oxford University Press on behalf of the Mind Association

Stable URL: <http://www.jstor.org/stable/2254605>

Accessed: 30-10-2017 15:41 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



Mind Association, Oxford University Press are collaborating with JSTOR to digitize, preserve and extend access to *Mind*

*Language and Nature*¹

NOAM CHOMSKY

1. Language as a natural object

I would like to discuss an approach to the mind that considers language and similar phenomena to be elements of the natural world, to be studied by ordinary methods of empirical inquiry. I will be using the terms “mind” and “mental” here with no metaphysical import. Thus I understand “mental” to be on a par with “chemical”, “optical”, or “electrical”. Certain phenomena, events, processes and states are informally called “chemical” etc., but no metaphysical divide is suggested thereby. The terms are used to select certain aspects of the world as a focus of inquiry. We do not seek to determine the true *criterion of the chemical*, or the *mark of the electrical*, or the *boundaries of the optical*. I will use “mental” the same way, with something like ordinary coverage, but no deeper implications. By “mind” I just mean the mental aspects of the world, with no more interest in sharpening the boundaries or finding a criterion than in other cases.

I’ll use the terms “linguistic” and “language” in much the same way. We focus attention on aspects of the world that fall under this informal rubric, and try to understand them better. In the course of doing so we may—and apparently do—develop a concept that more or less resembles the informal notion of “language”, and postulate that such objects are among the things in the world, alongside of complex molecules, electrical fields, the human visual system, and so on.

A naturalistic approach to linguistic and mental aspects of the world seeks to construct intelligible explanatory theories, taking as “real” what we are led to posit in this quest, and hoping for eventual unification with the “core” natural sciences: unification, not necessarily reduction. Large-scale reduction is rare in the history of the sciences. Commonly the more “fundamental” science has had to undergo radical revision for unification to proceed. The case of chemistry and physics is a recent example;

¹ §1 of this paper is based on the Homer Smith lecture at the New York University School of Medicine, May 16, 1994 and on the Jacobsen lecture, University of London, May 23, 1994. §2 is based on a lecture entitled “Linguistics from an Individualistic Perspective” delivered at the Centre for Philosophical Studies, King’s College London, May 24, 1994.

Pauling's account of the chemical bond unified the disciplines, but only after the quantum revolution in physics made these steps possible. The unification of much of biology with chemistry a few years later might be regarded as genuine reduction, but that is not common, and has no particular epistemological or other significance; "expansion" of physics to incorporate what was known about valence, the Periodic table, chemical weights, and so on is no less valid a form of unification. In the present case, the theories of language and mind that seem best established on naturalistic grounds attribute to the mind/brain computational properties of a kind that are well-understood, though not enough is known to explain how a structure constructed of cells can have such properties. That poses a unification problem, but of a familiar kind.

We do not know how eventual unification might proceed in this case, or if we have hit upon the right categories to seek to unify, or even if the question falls within our cognitive reach. We have no warrant simply to assume that mental properties are to be reduced to "neural network properties", to take a typical claim (Patricia Churchland 1994). Similar pronouncements have often proven false in other domains and are without any particular scientific merit in this case. If the thesis about neural networks is understood as a research proposal, well and good; we wait and see. If more is intended, rather serious questions arise.

As for the matter of cognitive reach, if humans are part of the natural world, not supernatural beings, then human intelligence has its scope and limits, determined by initial design. We can thus anticipate that certain questions will not fall within their cognitive reach, just as rats are unable to run mazes with numerical properties, lacking the appropriate concepts. Such questions we might call "mysteries-for-humans", just as some questions pose mysteries-for-rats. Among these mysteries may be questions we raise, and others we do not know how to formulate properly or at all. These truisms do not charge humans with "Feeble Intelligence". We do not condemn the human embryo as "feeble" because its genetic instructions are rich enough to enable it to become a human, hence to block other paths of development. Everyone would applaud if "questions shift status from Mysteries We Can Only Contemplate in Awe, to Tough Problems We Are Beginning to Crack" (Churchland 1994).² To demonstrate the shift for matters of traditional concern is no small order, and one may fairly ask whether the horizons remain as remote as ever, perhaps for reasons rooted in the human biological endowment.

Daniel Dennett argues that the notion of "epistemic boundedness", while "doctrinally convenient", is "rhetorically unstable", because

² The target of the derisive comments is McGinn (1991); McGinn points out the fallacy of the argument. See also McGinn (1993) and Chomsky (1975a).

“Chomsky and [Jerry] Fodor have hailed the capacity of the human brain to parse, and hence presumably understand, the official infinity of grammatical sentences of a natural language”, including those “that best express the solutions to the problems of free will or consciousness”, which he mistakenly claims I have declared “off-limits” (Dennett 1991). But even if the solutions can be formulated in human language—which has to be shown, not asserted—the argument is fallacious. First, as is well-known, expressions of natural language are often unparseable (not because of length, or complexity in some sense independent of the nature of the language faculty). Second, even if parsed and assigned an interpretation, they may be utterly incomprehensible; examples are all too easy to find.

The history of the advanced sciences offers some insights into the quest for unification. Take as a starting point the “mechanical philosophy” that reached its apogee in the 17th century: the idea that the world is a machine of the kind that could be constructed by a skilled craftsman. This conception of the world has its roots in common sense understanding, from which it drew the crucial assumption that objects can interact only through direct contact. As is familiar, Descartes argued that certain aspects of the world—crucially, the normal use of language—lie beyond the bounds of mechanism. To account for them, he postulated a new principle; in his framework, a second substance, whose essence is thought. The “unification problem” arose as a question about the interaction of body and mind. This metaphysical dualism was naturalistic in essence, using empirical evidence for factual theses about the world—wrong ones, but then, that is the rule.

The Cartesian theory collapsed soon afterwards, when Newton showed that terrestrial and planetary motion lie beyond the bounds of the mechanical philosophy—beyond what was understood to be body, or matter. What remained was a picture of the world that was “antimaterialist”, and that “relied heavily on spiritual forces”, as Jacob (1988) puts it.

Newton’s invocation of gravity was sharply condemned by leading scientists. E. J. Dijksterhuis points out that “the leaders of the true mechanistic philosophy regarded the theory of gravitation (to use the words of Boyle and Huygens) as a relapse into medieval conceptions that had been thought exploded, and as a kind of treason against the good cause of natural science” (1986, pp. 479ff). Newton’s “mysterious force” was a return to the dark ages from which scientists had “emancipated themselves”, “the scholastic physics of qualities and powers”, “animistic explanatory principles”, and the like, which admitted interaction without “direct contact”. It was as if “Newton had stated that the sun generates in the planets a quality which makes them describe ellipses”. In their corre-

spondence, Leibniz and Huygens condemn Newton for abandoning sound “mechanical principles” and reverting to mystical “sympathies and antipathies”, “immaterial and inexplicable qualities”. Newton seems to have agreed. The context of his famous comment that “I frame no hypotheses” was an expression of concern over his inability to “assign the cause of this power” of gravity, which so departs from “mechanical causes”. He therefore had to content himself with the conclusion “that gravity does really exist”, its laws explaining “all the motions of the celestial bodies, and of our sea”—though he regarded the principle he postulated as an “absurdity”. To the end of his life, Newton sought some “subtle spirit which pervades and lies hid in all gross bodies” that would account for interaction, electrical attraction and repulsion, the effect of light, sensation, and the way “members of animal bodies move at the command of the will”. Similar efforts continued for centuries (see Dijksterhuis 1986).

These concerns, at the origins of modern science, have something of the flavour of contemporary discussion of the “mind-body problem”. They also raise questions about what is at stake. Thomas Nagel observes that “the various attempts to carry out this apparently impossible task [of reducing mind to matter] and the arguments to show that they have failed, make up the history of the philosophy of mind during the past fifty years”. The hopeless task is to “complete the materialist world picture” by translating accounts of “mental phenomena” in terms of “a description that is either explicitly physical or uses only terms that can apply to what is entirely physical”, or perhaps gives “assertibility conditions” on “externally observable grounds” (Nagel 1993, p. 37). In an instructive review of a century of the philosophy of mind, Tyler Burge discusses the emergence of “naturalism” (“materialism”, “physicalism”) in the 1960s as “one of the few orthodoxies in American philosophy”: the view that there are no mental states (properties, etc.) “over and above ordinary physical entities, entities identifiable in the physical sciences or entities that common sense would regard as physical” (Burge 1992, pp. 31-2).

Such discussions assume, contrary to Newton and his contemporaries, that Newton remained within “the materialist world picture”; that would be true only if we understand “the materialist world picture” to be whatever science constructs, however it departs from “mechanical causes”. To put it differently, the discussions presuppose some antecedent understanding of what is physical or material, what are the physical entities. These terms had some sense within the mechanical philosophy, but what do they mean in a world based on Newton’s “mysterious force”, or still more mysterious notions of fields of force, curved space, infinite one-dimensional strings in ten-dimensional space, or whatever science concocts tomorrow? Lacking a concept of “matter” or “body” or “the

physical”, we have no coherent way to formulate issues related to the “mind-body problem”. These were real problems of science in the days of the mechanical philosophy. Since its demise, the sciences postulate whatever finds a place in intelligible explanatory theory, however offensive that may be to common sense. Only on unjustified dualistic assumptions can such qualms be raised specifically about the domain of the mental, not other aspects of the world.

The anti-materialism of the Newtonians soon became established. By mid-eighteenth century, Diderot’s materialist commitments were apparently a factor in his overwhelming rejection for membership in the Royal Society. Hume wrote that “Newton seemed to draw off the veil from some of the mysteries of nature”, but “he showed at the same time the imperfections of the mechanical philosophy; and thereby restored [Nature’s] ultimate secrets to that obscurity in which they ever did and ever will remain” (cited by Gay (1977, p. 130)).

That these secrets might remain in obscurity had sometimes been denied. Isaac Beekman, whom Jacob identifies as “the first mechanical philosopher of the Scientific Revolution”, was confident that “God had so constructed the whole of nature that our understanding ... may thoroughly penetrate all the things on earth” (Jacob 1988, p. 52). Similar theses are propounded with the same confidence today, notably by people who describe themselves as hard-headed scientific naturalists and who typically rephrase Beekman’s formula, replacing “God” by “natural selection”—with even less justification, because the *deus ex machina* is better defined in this case, so it is easy to see why the arguments fail.

Though Newton’s anti-materialism became scientific common sense, his qualms were not really put to rest. One expression of them was the belief that nature was unknowable. Another variant held that theoretical posits should be given only an operationalist interpretation. Lavoisier believed that “the number and nature of elements” is “an unsolvable problem, capable of an infinity of solutions none of which probably accord with Nature It seems extremely probable we know nothing at all about ... [the] ... indivisible atoms of which matter is composed”, and never will (cited by Brock 1992, p. 129). Boltzmann described his molecular theory of gases as nothing but a convenient analogy. Poincaré held that we have no reason to choose between ethereal-mechanical or electromagnetic theories of light, and that we accept the molecular theory of gases because we are familiar with the game of billiards. The chemist’s atoms were considered “theoretical, metaphysical entities”, William Brock observes; interpreted operationally, they provided a “conceptual basis for assigning relative elementary weights and for assigning molecular formulae”, and these instrumental devices were distinguished

from “a highly controversial physical atomism, which made claims concerning the ultimate mechanical nature of all substances”. Unification was only achieved with radical changes in physical atomism: Bohr’s model, quantum theory, and Pauling’s discoveries.³

The unification finally overcame what had seemed an unbridgeable divide, pre-Planck: “The chemist’s matter was discrete and discontinuous, the physicist’s energy continuous”, a “nebulous mathematical world of energy and electromagnetic waves...” (Brock 1992, p. 489).

In mid-19th century, the formulas analysing complex molecules were considered to be “merely classificatory symbols that summarized the observed course of a reaction”; the “ultimate nature of molecular groupings was unsolvable”, it was held, and “the actual arrangements of atoms within a molecule”, if that even means anything, is “never to be read” into the formulas. Kekulé, whose structural chemistry paved the way to eventual unification, doubted that “absolute constitutions of organic molecules could ever be given”; his models and analysis of valency were to have an instrumental interpretation only. Until the 1870s, Kekulé rejected the idea that the “rational formulae... actually represented the real arrangements of a molecule’s atoms”. As late as 1886, French schools were not permitted to teach atomic theory because it was a “mere hypothesis”, by decision of the Minister of Education, the well-known chemist Berthelot.

Forty years later, eminent scientists ridiculed as a conceptual absurdity the proposal of G.N. Lewis that “the atomic shells were mutually interpenetrable” so that an electron “may form part of the shell of two different atoms”—later “a cardinal principle of the new quantum mechanics” (Brock 1992, p. 476). It was “equivalent to saying that husband and wife, by having a total of two dollars in a joint account and each having six dollars in individual bank accounts, have got eight dollars apiece”, one objection ran; it was as if the electrons were “sitting around on dry goods boxes at every corner, ready to shake hands with... electrons in other atoms”, a distinguished Faraday lecturer commented with derision. America’s first Nobel Prize-winning chemist, Theodore Richards, dismissed talk about the real nature of chemical bonds as metaphysical “twaddle”. This was nothing more than “a very crude method of representing certain known facts about chemical reactions. A mode of represent[ation]” only. The rejection of that skepticism by Lewis and others paved the way to the eventual unification.

³ Cited by Brock (1992, pp. 165, 171). For reference to Boltzmann and Poincaré, see Chomsky (1986), which also cites John Heilbron’s unpublished PhD dissertation (University of California at Berkeley).

It is not hard to find contemporary counterparts in the discussion of the mind-body problem, whatever that is supposed to be. There is, I think, a good deal to learn from the history of the sciences since they abandoned common sense foundations, always with some uneasiness about just what they were doing. We should by now be able to accept that we can do no more than seek “best theories”, with no independent standard for evaluation apart from contribution to understanding, and hope for unification but with no advance doctrine about how, or whether, it can be achieved. As Michael Friedman puts the point, “the philosophers of the modern tradition”, from Descartes, “are not best understood as attempting to stand outside the new science so as to show, from some mysterious point outside of science itself, that our scientific knowledge somehow ‘mirrors’ an independently existing reality. Rather, [they] start from the *fact* of modern scientific knowledge as a fixed point, as it were. Their problem is not so much to justify this knowledge from some ‘higher’ standpoint as to articulate the new *philosophical* conceptions that are forced upon us by the new science.” In Kant’s words, mathematics and the science of nature stand in no need of philosophical inquiry for themselves, “but for the sake of another science: metaphysics” (Friedman 1993).

On this view, the natural sciences—whether the topic is the motion of the planets, the growth of an organism, or language and mind—are “first philosophy”. The idea is by now a commonplace with regard to physics; it is a rare philosopher who would scoff at its weird and counterintuitive principles as contrary to right thinking and therefore untenable. But this standpoint is commonly regarded as inapplicable to cognitive science, linguistics in particular. Somewhere between, there is a boundary. Within that boundary, science is self-justifying; the critical analyst seeks to learn about the criteria for rationality and justification from the study of scientific success. Beyond that boundary, everything changes; the critic applies independent criteria to sit in judgment over the theories advanced and the entities they postulate. This seems to be nothing more than a kind of “methodological dualism”, far more pernicious than the traditional metaphysical dualism, which was a scientific hypothesis, naturalistic in spirit. Abandoning this dualist stance, we pursue inquiry where it leads.

We also should be able now to adopt an attitude towards the mind-body problem formulated in the wake of Newton’s demolition of materialism and the “mechanical philosophy”: for example, by Joseph Priestley, whose conclusion was “not that all reduces to matter, but rather that the kind of matter on which the two-substance view is based does not exist”, and “with the altered concept of matter, the more traditional ways of posing the question of the nature of thought and of its relations to the brain do not fit. We have to think of a complex organized biological system with

properties the traditional doctrine would have called mental *and* physical” (Yolton 1983, p. 114).

In Priestley’s words, matter “is possessed of powers of attraction and repulsion” that act at a “real and in general an assignable distance from what we call the body itself”, properties that are “absolutely essential to [the] very nature” of matter. We thus overcome the naive belief that bodies (atoms aside) have inherent solidity and impenetrability, dismissing arguments based on “vulgar phraseology” and “vulgar apprehensions”, as in the quest for the *me* referred to in the phrase “my body”. With the Newtonian discoveries, matter “ought to rise in our esteem, as making a nearer approach to the nature of spiritual and immaterial beings”, the “odium [of] solidity, inertness, or sluggishness” having been removed. Matter is no more “incompatible with sensation and thought” than with attraction and repulsion. “The powers of sensation or perception and thought” are properties of “a certain organized system of matter”; properties “termed mental” are “the result (whether necessary or not) of such an organical structure as that of the brain”. It is as reasonable to believe “that the powers of sensation and thought are the necessary result of a particular organization, as that sound is the necessary result of a particular concussion of the air”. Thought in humans “is a property of the nervous system, or rather of the brain”.⁴

More cautiously, we may say that in appropriate circumstances *people* think, not their brains, which do not, though their brains provide the mechanisms of thought. I may do long division by a procedure I learned in school, but my brain doesn’t do long division even if it carries out the procedure. Similarly, I myself am not doing long division if I mechanically carry out instructions that are interpreted as the very algorithm I use, responding to inputs in some code in a Searle-style “arithmetic room”. Nothing follows about my brain’s executing an algorithm; likewise in the case of translation and understanding. *People* in certain situations understand a language; my brain no more understands English than my feet take a walk. It is a great leap from common sense intentional attributions to people, to such attributions to parts of people or to other objects. That move has been made far too easily, leading to extensive and it seems pointless debate over such alleged questions as whether machines can think: for example, as to “how one might *empirically* defend the claim that that a given (strange) object plays chess” (Haugeland 1979, p. 620), or determine whether some artifact or algorithm can translate Chinese, or reach for an object, or commit murder, or believe that it will rain. Many

⁴ The quotations from Priestley in this paragraph come from Passmore (1965), especially pp. 103ff. Similar conclusions had been drawn by La Mettrie a generation earlier, though on different grounds.

of these debates trace back to the classic paper by Alan Turing in which he proposed the Turing test for machine intelligence, but they fail to take note of his observation that “The original question, ‘Can machines think?’ I believe to be too meaningless to deserve discussion” (Turing 1950, p. 442). It is not a question of fact, but a matter of decision as to whether to adopt a certain metaphorical usage, as when we say (in English) that airplanes fly but comets do not—and as for space shuttles, choices differ. Similarly, submarines set sail but do not swim. There can be no sensible debate about such topics; or about machine intelligence, with the many familiar variants.

It is perhaps worth comparing contemporary debate with 17th-18th century discussion of similar topics. Then, too, many were intrigued by the capacities of artifacts, and debated whether humans might simply be devices of greater complexity and different design. But that debate was naturalistic in character, having to do with properties apparently not subsumed under the mechanical philosophy. Focusing on language use, Descartes and his followers, notably Géraud de Cordemoy, outlined experimental tests for “other minds”, holding that if some object passes the hardest experiments I can devise to test whether it expresses and interprets new thoughts as I do, it would be “unreasonable” to doubt that it has a mind like mine. This is ordinary science, on a par with a litmus test for acidity. The project of machine simulation was actively pursued, but understood as a way to find out something about the world. The great artificer Jacques de Vaucanson did not seek to fool his audience into believing that his mechanical duck was digesting food, but rather to learn something about living things by construction of models, as is standard in the sciences. Contemporary debate contrasts rather unfavourably with this tradition.⁵

Similar considerations hold with regard to the intentional terminology commonly used in describing what happens in the world. Thus we say that the asteroid is aiming toward the earth, the missile is rising toward the moon, the flower is turning toward the light, the bee is flying to the flower, the chimpanzee is reaching for the coconut, John is walking to his desk. Some future naturalistic theory might have something to say both about normal usage, and about the cases it seeks to address, two quite different topics. Neither inquiry would be bound by “vulgar phraseology [and] apprehensions”, just as we do not expect the theory of vision to deal with Clinton’s vision of the international market, or expect the theory of language to deal with the fact that Chinese is the language of Beijing and

⁵ See Marshall (1989); and Chomsky et al. (1993) for further comment; and for more extensive discussion, Chomsky (1966).

Hong Kong, though Romance is not the language of Bucharest and Rio de Janeiro—as a result of such factors as the stability of empires.

It would be misleading to say that we abandon the *theories* that the asteroid is aiming towards the earth, that the sun is setting and the heavens darkening, that the wave hit the beach and then receded, that the wind died and the waves disappeared, that people speak Chinese but not Romance, and so on, replacing them by better ones. Rather, the search for theoretical understanding pursues its own paths, leading to a completely different picture of the world, which neither vindicates nor eliminates our ordinary ways of talking and thinking. These we can come to appreciate, modify and enrich in many ways, though science is rarely a guide in areas of human significance. Naturalistic inquiry is a particular human enterprise that seeks a special kind of understanding, attainable for humans in some few domains when problems can be simplified enough. Meanwhile, we live our lives, facing as best we can problems of radically different kinds, far too rich in character for us to hope to be able to discern explanatory principles of any depth, if these even exist.⁶

The basic contention of Priestley and other 18th century figures seems uncontroversial: thought and language are properties of organized matter—in this case, mostly the brain, not the kidney or the foot. It is unclear why the conclusion should be resurrected centuries later as an audacious and innovative proposal: “the bold assertion that mental phenomena are entirely natural and caused by the neurophysiological activities of the brain” (Paul Churchland 1994), the hypothesis “that capacities of the human mind are in fact capacities of the human brain” (Patricia Churchland 1994); or that “consciousness is a higher-level or emergent property of the brain”, “as much of the natural biological order as ... photosynthesis, digestion, or mitosis” (Searle 1992), nor why Nagel should describe this last as the “metaphysical heart” of a “radical thesis” that “would be a major addition to the possible answers to the mind-body problem” if properly clarified (as he considers unlikely: Nagel 1993). Every year or two a book appears by some distinguished scientist with the “startling conclusion” or “astonishing hypothesis” that thought in humans “is a property of the nervous system, or rather of the brain”, the “necessary result of a particular organization” of matter, as Priestley put the matter long ago, in terms that seem close to truism—and as uninformative as truisms tend to be, since the brain sciences, despite important progress, are far from closing the gap to the problems posed by thought and language, or even to what is more or less understood about these topics.

⁶ For somewhat similar conclusions on different grounds, see Baker (1988) and Chastain (1988).

Here, we face typical problems of unification. “The variance of neural maps is not discrete or two-valued but rather continuous, fine-grained, and extensive”, Edelman (1992) writes, concluding that computational or connectionist theories of the mind must be wrong because of their discrete character. That is no more reasonable than the conclusion, a century ago, that chemistry must be wrong because it could not be unified with what we now know to be a far-too-impooverished physics; in particular, because “the chemist’s matter was discrete and discontinuous, the physicist’s energy continuous”.⁷ The disparity is real enough, but it is not, as Edelman sees it, a “crisis” for cognitive science; rather a unification problem, in which the chips fall where they may.

There is no problem of principle in devising systems that map continuous inputs into very specific discrete outputs; the “all-or-nothing” character of neural interaction is an example. Another illustration is given in a recent study that uses “a thermodynamic computer model to show that great regularity in the position of a subtle feature, a switch from six to four layers, can result from a slight discontinuity in the inputs to the lateral geniculate during development”, a “small perturbation” that “markedly affect[s] the overall organization of... a large structure”, one of many such examples, the author notes (Stryker 1994, p. 263). Whatever the empirical status of particular proposals, the problems of unification of discrete (computational or connectionist) and cellular theories have not been shown to be different in kind from others that have arisen throughout the course of science.

The current situation is that we have good and improving theories of some aspects of language and mind, but only rudimentary ideas about the relation of any of this to the brain. Consider a concrete example. Within computational theories of the language faculty of the brain, there is by now a fairly good understanding of distinctions among kinds of “deviance”—departure from one or another general principle of the language faculty. Recent work on electrical activity of the brain has found correlates to several of these categories of deviance, and a distinctive kind of electrophysiological response to syntactic vs. semantic violations.⁸ Still, the findings remain something of a curiosity, because there is no appropriate theory of electrical activity of the brain—no known reason, that is, why one should find these results, not others. The computational theories, in contrast, are more solidly based from the point of view of scientific nat-

⁷ See p. 6, above. For some comment on Edelman’s misinterpretation of the computational theories to which he alludes, and of the nature of semantics, in which he expects to find a solution to the “crisis”, see Chomsky et al. (1993).

⁸ See Neville et al. (1991), Hagoort et al. (1992), and Hagoort and Brown (1993).

uralism; the analysis of deviance, in particular, falls within an explanatory matrix of considerable scope.

A naturalistic approach to language and mind will seek to improve each approach, hoping for more meaningful unification. It is common to suppose that there is something deeply problematic in the theory that is more solidly established on naturalistic grounds, the “mental one”; and to worry about problems of “eliminationism” or “physicalism” that have yet to be formulated coherently. Furthermore, this dualist tendency not only dominates discussion and debate, but is virtually presupposed, a curious phenomenon of the history of thought that merits closer investigation.

Putting aside such tendencies, how would a naturalistic inquiry proceed? We begin with what we take to be natural objects, for example Jones. We are initially interested in particular aspects of Jones, the linguistic aspects. We find that some elements of Jones’s brain are dedicated to language—call them the *language faculty*. Other parts of the body may also have specific language-related design, and elements of the language faculty may be involved in other aspects of life, as we would expect of any biological organ. We set these matters to one side at first, keeping to the language faculty of the brain, clearly fundamental. There is good evidence that the language faculty has at least two different components: a “cognitive system” that stores information in some manner, and performance systems that make use of this information for articulation, perception, talking about the world, asking questions, telling jokes, and so on. The language faculty has an input receptive system and an output production system, but more than that: no one speaks only Japanese and understands only Swahili. These performance systems access a common body of information, which links them and provides them with instructions of some kind. The performance systems can be selectively impaired, perhaps severely so, while the cognitive system remains intact, and further dissociations have been discovered, revealing the kind of modular structure expected in any complex biological system.

Note that “modularity” here is not understood in the sense of Jerry Fodor’s interesting work, which keeps to input and output systems. The cognitive system of the language faculty is accessed by such systems, but is distinct from them. It may well be true that “psychological mechanisms” are “composed of independent and autonomous faculties like the perception of faces and of language” (Mehler and Dupoux 1994), but these “mental organs” do not appear to fit within the framework of modularity, as more narrowly construed. Similarly, David Marr’s influential ideas about levels of analysis do not apply here at all, contrary to much discussion, because he too is considering input-output systems; in this case, the mapping of retinal stimulations to some kind of internal image.

Jones's language faculty has an "initial state", fixed by genetic endowment. It is generally assumed that the performance systems are fully determined by the initial state—that any state changes are internally directed or are the result of extraneous factors such as injury, not exposure to one or another language. This is the simplest assumption, and it is not known to be false, though it may well be; adopting it, we attribute language-related differences in perception (say, our inability to perceive differences of aspiration as a Hindi speaker would) to differences in the phonetic aspects of the cognitive system, without having much faith in the assumption, though there is some evidence for it. (Under experimental conditions English speakers detect the Hindi contrasts that they do not "hear" in a linguistic context.) The performance systems may well be specialized for language. Even very young infants appear to have something like the adult phonetic system in place, perhaps a special refinement of a broader vertebrate category. Mehler and Dupoux (1994) propose the working hypothesis that "newborns are sensitive to *all* contrasts that can appear in *all* natural languages, and in exactly the same way as adults", with "learning by forgetting" under early exposure, so that before the child is a year old, the cognitive system has selected some subpart of the available potential.

On these simplifying assumptions about development, we look just at the cognitive system of the language faculty, its initial state, and its later states. Plainly, there are state changes that reflect experience: English is not Swahili, at least, not quite. A rational Martian scientist would probably find the variation rather superficial, concluding that there is one human language with minor variants. But the cognitive system of Jones's language faculty *is* modified in response to linguistic experience, changing state until it pretty much stabilizes, perhaps as early as six to eight years old, which would mean that later (nonlexical) changes that have been found, up to about puberty, are inner-directed.

Let us tentatively call a state of the cognitive system of Jones's language faculty a "language"—or to use a technical term, an "I-language", "I" to suggest "internal", "individual", since this is a strictly internalist, individualist approach to language, analogous in this respect to studies of the visual system.⁹ If the cognitive system of Jones's language faculty is in state *L*, we will say that Jones has the I-language *L*. An I-

⁹ Note that this interpretation of such studies differs from some that appear in the philosophical literature. The term "I-language" was introduced to overcome misunderstanding engendered by the systematic ambiguity of the term "grammar", used both to refer to an I-language and to the linguist's theory of it. Thus Jones's knowledge of his I-language (grammar, in one sense) is nothing like some linguist's (partial) knowledge.

language is something like “a way of speaking”, one traditional notion of language.

Despite some similarity to standard locutions, however, the terminology here is different, as we expect even in the earliest stages of naturalistic inquiry. The languages of the world describe such matters in various ways. In English, we say that Jones *knows* his language; others say that he speaks it, or speaks with it, and so on, and terms for something like language vary, though I know of no serious cross-cultural study. These topics are of interest for natural language semantics, and other branches of naturalistic inquiry that seek to determine how cognitive systems, including language, yield what is sometimes called “folk science”. We speak of flowers turning toward the sun, the heavens darkening, apples falling to the ground, people having beliefs and speaking languages, and so on; our ways of thinking and understanding, and our intuitive ideas about how the world is constituted, may or may not relate directly to such locutions. The elements of folk science derive from our biological endowment, taking particular forms under varying cultural conditions. There is evidence that young children attribute beliefs and plans to others well before they have terms to describe this, and the same may be true of adults generally, though most languages, it is reported, do not have terms corresponding to the English “belief”. These are serious inquiries, not to be undertaken casually; our intuitions about them provide some evidence, but nothing more than that. Furthermore, whatever may be learned about folk science will have no relevance to the pursuit of naturalistic inquiry into the topics that folk science addresses in its own way, a conclusion taken to be a truism in the study of what is called “the physical world” but considered controversial or false (on dubious grounds, I think) in the study of the mental aspects of the world.

So far I have kept to Jones, his brain, its language faculty, and some of its components, all natural objects. Turning to Smith, we discover that the initial state of his language faculty is virtually identical; given Jones’s experience, he would have Jones’s language. That seems to be true across the species, meaning that the initial state is a species property, to a very good first approximation. If so, the *human language faculty* and the (I-)languages that are manifestations of it qualify as natural objects.

If Jones has the language *L*, he knows many things: for example, that *house* rhymes with *mouse* and that *brown house* consists of two words in the formal relation of assonance, and is used to refer to a structure designed and used for certain purposes and with a brown exterior. We would like to find out how Jones knows such things. It seems to work something like this.

The I-language consists of a computational procedure and a lexicon. The lexicon is a collection of items, each a complex of properties (called “features”), such as the property “bilabial stop” or “artifact”. The computational procedure selects items from the lexicon and forms an expression, a more complex array of such features. There is reason to believe that the computational system is virtually invariant; there is some variation at the parts closely related to perception and articulation, not surprisingly, since it is here that data are available to the child acquiring language—a process better described as “growth” than as “learning”, in my opinion. That aside, language variation appears to reside in the lexicon. One aspect is “Saussurean arbitrariness”, the arbitrary links between concepts and sounds: the genetic program does not determine whether *tree*, the concept, is associated with the sounds “tree” or “Baum”. The linkage of concept and sound can be acquired on minimal evidence, so variation here is not surprising. But the possible sounds are narrowly constrained, and the concepts may be virtually fixed. It is hard to imagine otherwise, given the rate of lexical acquisition—about a word an hour from ages two to eight, with lexical items typically acquired on a single exposure, in highly ambiguous circumstances, but understood in delicate and extraordinary complexity that goes vastly beyond what is recorded in the most comprehensive dictionary—which, like the most comprehensive traditional grammar, merely gives hints that suffice for people who basically know the answers, largely innately.

Beyond such factors, variation may be limited to formal aspects of language—case of nouns, verbal inflection, and so on. Even here, variation may be slight. On the surface, English appears to differ sharply from German, Latin, Greek or Sanskrit in richness of inflection; Chinese even more so. But there is evidence that the languages have basically the same inflectional systems, differing only in the way formal elements are accessed by the part of the computational procedure that provides instructions to articulatory and perceptual organs. The mental computation seems otherwise identical, yielding indirect effects of inflectional structure that are observable, even if the inflections themselves are not heard in speech. That may well be the basis of language variation, in large measure. Small changes in the way a system functions may, of course, yield what appears to be great phenomenal variety.

The computational procedure has properties that may be unique to it, in substantial part. It is also “austere”, with no access to many of the properties of other cognitive systems. For example, it seems to have no “counters”. It registers adjacency; thus every other syllable could have some property (say, stress). But it cannot use the notion *three*. There are no known phonological systems in which something happens every third

syllable, for example; and syntax seems to observe a property of “structure-dependence”, unable to make use of linear and arithmetical properties that are much simpler to implement outside the language faculty.

Recent experimental work by Neil Smith and his colleagues (1993) bears on this matter. They have been studying a person—called “Christopher”—who seems to have an intact language faculty but severe cognitive deficits, an example of the kind of modularity of mental architecture that has been found repeatedly. Christopher had mastered some 16 languages, and can translate from them to English. The experiments involved Christopher and a control group. Both were taught Berber and an invented system designed to violate principles of language. As expected, Christopher learned Berber easily, but lacking other cognitive capacities, could do little with the invented system. The control group made some progress on the invented system, apparently treating it as a puzzle. But there were some extremely simple rules they did not discover: for example, the rule that placed an emphatic marker on the third word of a sentence. It seems that the “austerity” of the language faculty sufficed to bar discovery of a simple structure-independent rule, within a linguistic context. Our use of language of course involves numbers; we can understand and identify sonnets, for example. It also involves inference, though it seems that the computational procedure is too austere to use these resources either. The language faculty is both very rich and very impoverished, as any biological system is expected to be: capable of a high-level of achievement in specific domains, and correspondingly unable to deal with problems that lie outside them. As noted earlier, we should expect that to be true of all our faculties, including what might be called the “science-forming faculty”, the particular collection of qualities and abilities we use in conducting naturalistic inquiry.

Though highly specialized, the language faculty is not tied to specific sensory modalities, contrary to what was assumed not long ago. Thus, the sign language of the deaf is structurally much like spoken language, and the course of acquisition is very similar. Large-scale sensory deficit seems to have limited effect on language acquisition. Blind children acquire language as the sighted do, even colour terms and words for visual experience like “see” and “look”. There are people who have achieved close to normal linguistic competence with no sensory input beyond what can be gained by placing one’s hand on another person’s face and throat. The analytic mechanisms of the language faculty seem to be triggered in much the same ways whether the input is auditory, visual, even tactual, and, seem to be localized in the same brain areas.

These examples of impoverished input indicate the richness of innate endowment—though normal language acquisition is remarkable enough,

as even lexical access shows, not only because of its rapidity and the intricacy of result. Thus, very young children can determine the meaning of a nonsense word from syntactic information in a sentence far more complex than any they can produce (Gleitman 1990).

A plausible assumption today is that the principles of language are fixed and innate, and that variation is restricted in the manner indicated. Each language, then, is (virtually) determined by a choice of values for lexical parameters: with one array of choices, we should be able to deduce Hungarian; with another, Yoruba. This principles-and-parameters approach offers a way to resolve a fundamental tension that arose at the very outset of generative grammar. As soon as the first attempts were made to provide actual descriptions of languages, 40 years ago, it was discovered that the intricacy of structure is far beyond anything that had been imagined, that traditional descriptions of form and meaning merely skimmed the surface, while structuralist ones were almost irrelevant. Furthermore, the apparent variability of languages explodes as soon as one attends to facts that had been tacitly assigned to the unanalysed “intelligence of the reader”. To attain “descriptive adequacy”, it seemed necessary to give very intricate accounts, specific to particular languages, indeed to particular constructions in particular languages: complex rules for relative clauses in English, for example. But it was obvious that nothing of the sort could be true. The conditions of language acquisition make it plain that the process must be largely inner-directed, as in other aspects of growth, which means that all languages must be close to identical, largely fixed by initial state. The major recent research effort has been guided by this tension, pursuing the natural approach: abstracting from the welter of descriptive complexity certain general principles governing computation that would allow the rules of a particular language to be given in very simple forms, with restricted variety.

Efforts to resolve the tension in this way led finally to the principles-and-parameters approach just outlined. It is more a bold hypothesis than a specific theory, though parts of the picture are being filled in, and new theoretical ideas are leading to a vast expansion in relevant empirical materials from typologically diverse languages.

These ideas constitute a radical departure from a rich tradition of some 2500 years. If correct, they show not only that languages are cast to very much the same mould, with a near invariant computational procedure and only restricted lexical variation, but also that there are no rules or constructions in anything like the traditional sense, which was carried over to early generative grammar: no rules for formation of relative clauses in English, for example. Rather, the traditional constructions—verb phrase,

relative clause, passive, etc.—are taxonomic artifacts, their properties resulting from the interaction of far more general principles.

The principles-and-parameters approach dissociates two notions that fell together under the concept of I-language: there is a clear conceptual distinction between the state of the language faculty, on the one hand, and an instantiation of the initial state with parameters fixed, on the other. Apart from miracles, the objects so identified will always differ empirically. The actual state of one's language faculty is the result of interaction of a great many factors, only some of which are relevant to inquiry into the nature of language. On more theory-internal grounds, then, we take an I-language to be an instantiation of the initial state, idealizing from actual states of the language faculty. As elsewhere in naturalistic inquiry, the term "idealization" is somewhat misleading: it is the procedure we follow in attempting to discover reality, the real principles of nature. Only in the study of mental aspects of the world is this considered illegitimate, another example of pernicious dualism that should be overcome.

Progress along these lines has opened up new questions, notably, the question to what extent the principles themselves can be reduced to deeper and natural properties of computation. To what extent, that is, is language "perfect", relying on natural optimality conditions and very simple relations? One theory holds that, apart from the phonetic features that are accessed by articulatory-perceptual systems, the properties of an expression that enter into language use are completely drawn from the lexicon: the computation organizes these in very restricted ways, but adds no further features. That is a considerable simplification of earlier assumptions, which would, if correct, require considerable rethinking of the "interface" between the language faculty and other systems of the mind. Another recent theory, proposed in essence by Richard Kayne, is that there is no parametric variation in temporal order. Rather, order is a reflex of structural properties determined in the course of computation: all languages are of the basic form subject-verb-object. Other recent work seeks to show that possible expressions that would be interpretable at the interface, if formed, are barred by the fact that other computations with the same lexical resources are more economical.¹⁰

On such assumptions, we expect that languages are "learnable", because there is little to learn, but are in part "unusable", one reason being that global economy conditions may yield high levels of computational complexity. That languages are "learnable" would be a surprising empirical discovery; there is no general biological or other reason why languages made available by the language faculty should be fully accessible, as they will be if languages are fixed by the setting of simple param-

¹⁰ On these matters, see Chomsky (1993b, 1994); and sources cited therein.

eters. The conclusion that languages are partially unusable, however, is not at all surprising. It has long been known that performance systems often “fail”, meaning that they provide an analysis that differs from that determined by the cognitive system (the I-language). Many categories of expressions have been studied that pose structural problems for interpretation: multiple-embedding, so-called “garden path sentences”, and others. Even simple concepts may pose hard problems of interpretation; words that involve quantifiers or negation, for example. Such expressions as “I missed (not) seeing you last summer” (meaning I expected to see you but didn’t) cause endless confusion. Sometimes confusion is even codified, as in the idiom “near miss”, which means “nearly a hit”, not “nearly a miss” (analogous to “near accident”).

The belief that parsing is “easy and quick”, in one familiar formula, and that the theory of language design must accommodate this fact, is erroneous; it is not a fact. The problem, however, is to show that those parts of language that are usable are properly determined by the theories of computation and performance, no small matter.

Questions of this sort bring us to the borders of current inquiry. These are questions of a new order of depth, hence of interest, in the study of language and mind.

Other questions have to do with interface properties: how do the performance systems make use of expressions generated by the I-language? Some features of these expressions provide instructions only to articulatory and perceptual systems; thus one element of a linguistic expression is its *phonetic form*. It is generally assumed that these instructions are common to both articulation and perception, which is not at all obvious, hence interesting if true. Other properties of the expression provide instructions only for conceptual-intentional systems; this element of the expression is usually called *logical form*, but in a technical sense that differs from other usages; call it *LF* to avoid misunderstanding. Again, it is assumed that there is only one such array of instructions, and that it is dissociated from phonetic form. These assumptions are even more implausible, hence if true, very interesting discoveries.

On such assumptions, the computational procedure maps an array of lexical choices into a pair of symbolic objects, phonetic form and LF, and does so in a way that is optimal, from a certain point of view. The elements of these symbolic objects can be called “phonetic” and “semantic” features, respectively, but we should bear in mind that all of this is pure syntax, completely internalist, the study of mental representations and computations, much like the inquiry into how the image of a cube rotating in space is determined from retinal stimulations, or imagined. We may take the semantic features *S* of an expression *E* to be its *meaning* and the

phonetic features *P* to be its *sound*; *E* means *S* in something like the sense of the corresponding English word, and *E sounds P* in a similar sense, *S* and *P* providing the relevant information for the performance systems.

An expression such as “I painted my house brown” is accessed by performance systems that interpret it, on the receptive side, and articulate it while typically using it for one or another speech act, on the productive side. How is that done? The articulatory-perceptual aspects have been intensively studied, but these matters are still poorly understood. At the conceptual-intentional interface the problems are even more obscure, and may well fall beyond human naturalistic inquiry in crucial respects.

Perhaps the weakest plausible assumption about the LF interface is that the semantic properties of the expression focus attention on selected aspects of the world as it is taken to be by other cognitive systems, and provide intricate and highly specialized perspectives from which to view them, crucially involving human interests and concerns even in the simplest cases. In the case of “I painted my house brown”, the semantic features impose an analysis in terms of specific properties of intended design and use, a designated exterior, and indeed far more intricacy. If I paint my house brown, it has a brown exterior, but I can paint my house brown *on the inside*. The exterior-interior dimension has a marked and unmarked option; if neither is indicated, the exterior is understood. That is a typical property of the lexicon; if I say Jones climbed the mountain, I mean that he was (generally) going up, but I can say that he climbed *down the mountain*, using the marked option. If I am inside my house, I can clean it, affecting only the interior, but I cannot see it, unless an exterior surface is visible (through a window, for example). And I certainly cannot be near my house if I am inside it, even though it is a surface, in the unmarked case. Similarly, a geometrical cube is just a surface, but if we are using natural language, a point inside the cube cannot be near it. These properties hold quite generally: of boxes, igloos, airplanes, mountains, and so on. If I look through a tunnel in a mountain and see a lighted cave within, I do not see the mountain; only if I see its exterior surface (say, from inside the cave, looking through the tunnel at a mirror outside that reflects the surface). The same is true of impossible objects. If I tell you that I painted a spherical cube brown, you take its exterior to be brown in the unmarked case, and if I am inside it, you know I am not near it. And so on, to intricacy that has been far underestimated, and that poses problems of “poverty of stimulus” so extreme that knowledge of language in these regards too can only be assumed to be in substantial measure innately determined, hence virtually uniform among languages, much as we assume without discussion or understanding for other aspects of growth and development.

Quite typically, words offer conflicting perspectives. A city is both concrete and abstract, both animate and inanimate: Los Angeles may be pondering its fate grimly, fearing destruction by another earthquake or administrative decision. London is not a place. Rather, it is *at* a place, though it is not the things at that place, which could be radically changed or moved, leaving London intact. London could be destroyed and rebuilt, perhaps after millennia, still being London; Carthage could be rebuilt today, just as Tom Jones, though perfectly concrete, could be reincarnated as an insect or turned by a witch into a frog, awaiting the princess's kiss, but Tom Jones all along—concepts available to young children without instruction or relevant experience.

The abstract character of London is crucial to its individuation. If London is reduced to dust, *it*—that is, London—can be *re*-built elsewhere and be the *same* city, London. If my house is reduced to dust, it (my house) can be rebuilt elsewhere, but it won't be the same house. If the motor of my car is reduced to dust, it cannot be rebuilt, though if only partially damaged, it *can*. Pronouns involve dependency of reference, but not necessarily to the same thing; and both referential dependence and the narrower notion of sameness involve roles in a highly intricate space of human interests and concerns. Judgments can be rather delicate, involving factors that have barely been explored.

There are plenty of real examples illustrating such properties of terms of natural language. We have no problem understanding a report in the daily press about the unfortunate town of Chelsea, which is “preparing to move” (viewed as animate), with some residents opposed because “by moving the town, it will take the spirit out of it”, while others counter that “unless Chelsea moves, floods will eventually kill it”. There is a city called both “Jerusalem” and “al-Quds”, much as London is called “London” and “Londres”. What is this city? Its site is a matter of no small contention, even of UN Security Council resolutions. The government that claims it as its capital city has been considering plans to move al-Quds, while leaving Jerusalem in place. The chairman of the development authority explained that “We need to find a capital for the Palestinians, we have to find a site for al-Quds”—somewhere northeast of Jerusalem. The proposal is perfectly intelligible, which is why it greatly troubles people concerned about al-Quds. The discussion would pose puzzles of a kind familiar in the philosophical literature, even more so if the proposal were implemented—if, failing to observe some of Wittgenstein's good advice, we were to suppose that words like “London” or “Jerusalem” refer to things in the world in some public language, and were to try to sharpen meanings and ideas for conditions under which the presuppositions of normal use do not hold.

Even the status of (nameable) thing, perhaps the most elementary concept we have, depends crucially on such intricate matters as acts of human will, again something understood without relevant experience, determined by intrinsic properties of the language faculty and others. A collection of sticks in the ground could be a (discontinuous) thing—say, a picket fence, a barrier, a work of art. But the same sticks in the ground are not a thing if left there by a forest fire.¹¹

The matter of space-time continuity has no particular relevance to these issues, contrary to what is sometimes assumed (see Putnam 1993). Discontinuity of things is not at all in question; the United States is discontinuous in space, though it has become a nameable thing (shifting over time from plural to singular usage); an utterance or theatrical performance may be discontinuous in time. As just noted, discontinuous objects are readily understood as nameable things, within a proper matrix of human interests. Whether a city is understood within “folk science” as a (possibly) discontinuous four-dimensional object is a question of fact. The assumption that it is, or that semantic theory should say that it is, requires quite unnatural interpretations of such terms as “move (Chelsea)”, “the former (Chelsea)”, etc., issues easily overlooked, given a narrow concentration on object-reference. The properties and perspectives involved in individuating cities, houses, and the like remain to be discovered and explained, independent of the question of continuity.

Substances reveal the same kinds of special mental design. Take the term “water”, in the sense proposed by Hilary Putnam: as coextensive with “H₂O give or take certain impurities” (Putnam 1993, alluding to his 1975). Even in such a usage, with its questionable invocation of natural science, we find that whether something is water depends on special human interests and concerns, again in ways understood without relevant experience; the term “impurities” covers some difficult terrain. Suppose cup₁ is filled from the tap. It is a cup of water, but if a tea bag is dipped into it, that is no longer the case. It is now a cup of tea, something different. Suppose cup₂ is filled from a tap connected to a reservoir in which tea has been dumped (say, as a new kind of purifier). What is in cup₂ is water, not tea, even if a chemist could not distinguish it from the present contents of cup₁. The cups contain the same thing from one point of view, different things from another; but in either case cup₂ contains only water and cup₁ only tea. In cup₂, the tea is an “impurity” in Putnam’s sense, in cup₁ it is not, and we do not have water at all (except in the sense that milk is mostly water, or a person for that matter). If cup₃ contains pure H₂O into which a

¹¹ On such matters, and their significance for Quinean and similar theories of learning, see Chomsky (1975a, p. 203).

tea bag has been dipped, it is tea, not water, though it could have a higher concentration of H_2O molecules than what comes from the tap or is drawn from a river. Note that this is a particularly simple case, unlike its classic counterparts “earth”, “air”, “fire”, among many others.

Proceeding beyond the simplest cases, intricacies mount. I can paint the door to the kitchen brown, so it is plainly concrete; but I can walk through the door to the kitchen, switching figure and ground. The baby can finish the bottle and then break it, switching contents and container with fixed intended reference. There is interesting work by James Pustejovsky studying regularities in such systems, drawing on ideas of Julius Moravcsik (1975, 1990), Aristotelian in origin.¹² As we move on to words with more complex relational properties and the structures in which they appear, we find that interpretation is guided in fine detail by the cognitive system in ways that we expect to vary little because they are so remote from possible experience.

Neurologist Rodolfo Llinás (1987) puts the matter well when he describes perception as “a dream modulated by sensory input”, the mind being a “computational state of the brain generated by the interaction between the external world and an internal set of reference frames”. But the internal frames that shape the dreams are far more intricate and intriguing than often assumed, even at the level of the lexicon, still more so when we turn to expressions formed by the computational procedures.

Spelling out the properties of expressions, we learn more about the instructions at the LF (“semantic”) interface, which are interpreted in some manner to think and talk about the world, along with much else. Important and obscure questions still lie beyond: in what respects, for example, do these properties belong to the language faculty as distinct from other faculties of mind to which it is linked? How do lexical resources relate to belief systems, for example? Such questions remain within the domain of what people know, not what they do. Answers to them would still leave us far short of understanding how the resources of the cognitive systems are put to use. From this welter of issues it is hard to see how to extricate very much that might be subjected to naturalistic inquiry.¹³

Note that the properties of such words as “house”, “door”, “London”, “water”, and so on do not indicate that people have contradictory or otherwise perplexing beliefs. There is no temptation to draw any such conclusion, if we drop the empirical assumption that words pick out things, apart from particular usages, which they constrain in highly intricate ways.

¹² See also Pustejovsky (1993b), and other papers in Pustejovsky (1993a); and also Chomsky (1975a).

¹³ For some comment, see Chomsky (1993a).

Should we assume that expressions pick out things, intrinsically? More generally, should the “weakest assumptions” about the interface relations and the way they enter into thought and action be supplemented to include relations that hold between certain expressions and external things? That is commonly assumed, though we have to take care to distinguish two variants: (1) things in the world, or (2) things in some kind of mental model, discourse representation, and the like.¹⁴ If the latter, then the study is again internalist, a form of syntax. Suppose the former, and continue to assume that there are two interface levels, phonetic form and LF.

Suppose we postulate that corresponding to an element α of phonetic form there is an external object $*\alpha$ that α selects as its *phonetic value*; thus the element [ba] in Jones’s I-language picks out some entity $*[ba]$, “shared” with Smith if there is a counterpart in his I-language. Communication could then be described in terms of such (partially) shared entities, which are easy enough to construct: take $*\alpha$ to be the singleton set $\{\alpha\}$, or $\{3, \alpha\}$; or if one wants a more realistic feel, some construct based on motions of molecules. With sufficient heroism, one could defend such a view, though no one does, because it’s clear that we are just spinning wheels.

The same can be done at the LF interface. Suppose that α is constructed by the computational system from one or more lexical choices, where α is an LF representation or some further syntactic object computed from it (an expression in some formal language, some kind of mental model, etc.). We could then posit an object $*\alpha$ as its *semantic value*, external to the I-language, perhaps shared by Jones and Smith. Again, $*\alpha$ could be some arbitrary construction to which we assign the desired properties, or given a touch of realism in a variety of ways. We could then construct truth theories, and develop an account of communication in terms of shared entities—often of a very strange sort, to be sure. As in the case of any theoretical proposal that introduces new entities and principles, what has to be shown is that this one is justified in the usual empirical terms (explanatory power, etc.).

A good part of contemporary philosophy of language is concerned with analysing alleged relations between expressions and things, often exploring intuitions about the technical notions “denote”, “refer”, “true of”, etc., said to hold between expressions and something else. But there can be no intuitions about these notions, just as there can be none about “angular velocity” or “protein”. These are technical terms of philosophical discourse with a stipulated sense that has no counterpart in ordinary language—which is why Frege had to provide a new technical meaning

¹⁴ I put aside, here and below, the further assumption that these relations hold of objects in a public language. This notion is unknown to empirical inquiry, and raises what seem to be irresolvable problems, so far unaddressed. For some recent discussion, see Chomsky (1993a) and Chomsky et al. (1993).

for “Bedeutung”, for example. If we re-run the thought experiments with ordinary terms, judgments seem to collapse, or rather, to become so interest-relative as to yield no meaningful results.

Without pursuing the matter here, it is not at all clear that the theory of natural language and its use involves relations of “denotation”, “true of”, etc., in anything like the sense of the technical theory of meaning.

It is sometimes claimed that such technical notions are required to account for communication or for consideration of truth and falsity. The former belief is groundless. The latter also seems incorrect. Simply consider the ordinary language terms with which this discussion began: “language” and “mind”. Consider two statements about language and mind:

- (1) Chinese is the language of Beijing and Hong Kong, but not Melbourne.
- (2) The mind is its own place, and in itself can make a Heaven of Hell, a Hell of Heaven.

The first is true, but “Chinese” surely has no real world denotatum, in the technical sense, nor need one believe that it does to assign truth value. If we are convinced by Milton’s argument, we will agree that the second sentence is true, but without committing ourselves to the belief that any of the subject, the pronoun, or the reflexive (or the other noun phrases) refer either to something in the natural world or in some obscure mental world. At least, there is no compulsion to succumb to such temptations, for reasons put forth in the 18th century critique of the theory of ideas, much enriched in modern ordinary language philosophy. Such properties are typical of the words of natural language. This is not to deny that such statements can be made with referential intentions, but these are of a far more intricate nature.

In any event, there seems to be no special connection between attribution of truth or falsity and some notion of reference or denotation, in anything like the sense of technical discourse.

Consider in contrast another term I have used: *I-language*, which figures in such statements as:

- (3) I-language has a head parameter.

This statement is false if Kayne’s theory is correct, perhaps true if it is not. In this case, it makes sense to say that the term “I-language” has a real world denotatum, or at least is intended to. The statement belongs to the same kind of discourse as statements about H₂O, acids and bases, the specification of proteins by genes, etc. The sentences do not really belong to natural language; they contain technical terms, such as “I-language”, introduced in a quite different way. As the disciplines progress, they

depart still further from the common sense and ordinary language origins of inquiry.

It is reasonable to suppose that in the course of such inquiry, we attempt to construct systems in which well-constructed symbolic objects are intended to pick out objects in the world: molecules, I-languages, and so on. These symbolic systems may be called “languages”, but that is just a metaphor. They typically do not have properties of natural language, are acquired and used in a completely different fashion, and are not instantiations of the initial state of the language faculty. We may articulate symbolic objects of these systems with the phonetics of our language and borrow constructions of our language in using them, even when they contain terms that are invented or based on languages we do not know (“eigenvector”, “homo sapiens”), but all of that is irrelevant. The systems may depart in arbitrary ways from natural language, using calculus, chemical notations and diagrams, or whatever.

These symbolic systems may well aim towards to the Fregean ideal. According to this approach, there is a “common, public language” with formulas or signals that express shared thoughts. The “language” has a syntax, namely, a class of well-formed formulas; there is no “right answer” to the question of how that set is generated. It also has a semantics, based on the technical notion of *Bedeutung*, a relation between symbols and things. Perhaps one property of the science-forming faculty of the human mind is that it aims to construct Fregean systems. But if so, that will tell us nothing about natural language. Here there is no counterpart to the notion “common” or “public” language. The syntax is radically different. There a real answer to the question of what is the “right generative procedure”; I-languages are functions regarded in intension. And there appears to be no notion of “well-formed formula” in the sense used, for example, by Quine in his discussions of extensional equivalence and indeterminacy of translation, or by many linguists, psychologists, philosophers, and others who have been concerned about generative capacity, decidability of well-formedness, reduction to context-free grammars, excess strength of certain theories, and other problems that cannot even be formulated for natural language, as far as we know (Cf. Chomsky 1980, 1986).

As for semantics, insofar as we understand language use, the argument for a reference-based semantics (apart from an internalist syntactic version) seems to me weak. It is possible that natural language has only syntax and pragmatics; it has a “semantics” only in the sense of “the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation, is actually put to use in a speech community”, to quote the earliest formulation in generative

grammar 40 years ago, influenced by Wittgenstein, Austin and others (Chomsky (1975b, Preface), and Chomsky (1957, pp. 102–3). In this view, natural language consists of internalist computations and performance systems that access them along with much other information and belief, carrying out their instructions in particular ways to enable us to talk and communicate, among other things. There will be no provision for what Scott Soames calls “the central semantic fact about language, ... that it is used to represent the world”, because it is not assumed that language is used to represent the world, in the intended sense (Soames 1989, cited by Smith 1992 as the core issue for philosophers of language).

Before turning to more detailed issues relating to the internalist perspective on language, let me mention some limits. Some have already been suggested: general issues of intentionality, including those of language use, cannot reasonably be assumed to fall within naturalistic inquiry. The matter can be further clarified by returning to Cartesian dualism, the scientific hypothesis that sought to capture, in particular, the apparent fact that normal language use lies beyond the bounds of any possible machine. The Cartesian framework was undermined by the discovery that even the behavior of inorganic matter lies beyond these bounds. But the arguments can be reconstructed, though now without metaphysical implications, since the concept of matter has disappeared. So restated, they still seem to pose a complete mystery. They are, for example, unaffected by the transition from the complex artifacts that intrigued the Cartesians to today’s computers, and the brain sciences shed little light on them.

Possibly, as some believe, these problems are unreal. Possibly they are real but we have not hit upon the way to approach them. Possibly that way, whatever it is, lies outside our cognitive capacities, beyond the reach of the science-forming faculty. That should not surprise us, if true, at least if we are willing to entertain the idea that humans are part of the natural world, with rich scope and corresponding limits, facing problems that they might hope to solve and mysteries that lie beyond their reach, “ultimate secrets of nature” that “ever will remain” in “obscurity” as Hume supposed, echoing some of Descartes’s own speculations.

2. *Language from an internalist perspective*

I want to distinguish an *internalist* from a *naturalistic* approach. By the latter I mean just the attempt to study humans as we do anything else in the natural world, as discussed in §1. Internalist naturalistic inquiry seeks to understand the internal states of an organism. Naturalistic study is of

course not limited to such bounds; internalist inquiry into a planet or an ant does not pre-empt or preclude the study of the solar system or an ant community. Non-internalist studies of humans can take many forms: as phases in an oxygen-to-carbon dioxide cycle or gene transmission, as farmers or gourmets, as participants in associations and communities, with their power structures, doctrinal systems, cultural practices, and so on. Internalist studies are commonly presupposed in others with broader range, but it should be obvious that the legitimacy of one or another kind of inquiry does not arise.

To clarify further, I am keeping here to the quest for theoretical understanding, the specific kind of inquiry that seeks to account for some aspects of the world on the basis of usually hidden structures and explanatory principles. Someone committed to naturalistic inquiry can consistently believe that we learn more of human interest about how people think and feel and act by studying history or reading novels than from all of naturalistic inquiry. Outside of narrow domains, naturalistic inquiry has proven shallow or hopeless, and perhaps always will, perhaps for reasons having to do with our cognitive nature.

The aspects of the world that concern me here I will call its mental and linguistic aspects, using the terms innocuously—in the manner of “chemical”, “electrical”, or “optical”—to select a complex of phenomena, events, processes and so on that seem have a certain unity and coherence. By “mind”, I mean the mental aspects of the world. In none of these cases is there any need for antecedent clarity, nor any reason to believe that the categories will survive naturalistic inquiry where it can make some progress.

By “naturalism” I mean “methodological naturalism”, counterposed to “methodological dualism”: the doctrine that in the quest for theoretical understanding, language and mind are to be studied in some manner other than the ways we investigate natural objects, as a matter of principle. As discussed in §1, this is a doctrine that few may espouse, but that dominates much practice. See also Chomsky 1992, 1993a, forthcoming.

One branch of naturalistic inquiry studies common sense understanding. Here we are concerned with how people interpret object constancy, the nature and causes of motion, thought and action, and so on (“folk science”, in one of the senses of the term). Perhaps the right way to describe this is in terms of beliefs about the constituents of the world (call them “entities”) and their organization, interaction, and origins. Assume so. It is an open question whether, and if so how, the conceptual resources of folk science relate to those involved in the reflective and self-conscious inquiry found in every known culture (“early science”), and to the partic-

ular enterprise we call “natural science”. For convenience, let’s refer to the study of all such matters as “ethnoscience”.

It is also an open question how the conceptual resources that enter into these cognitive systems relate to the semantic (including lexical) resources of the language faculty. Do people attribute beliefs if they speak languages that have no such term, the great majority, it appears? Can someone lacking the terms recognize *savoir faire*, *Schadenfreude*, *machismo*, or whatever is expressed by the countless locutions that challenge translators? If I say that one of the things that concerns me is the average man and his foibles, or Joe Sixpack’s priorities, or the inner track that Raytheon has on the latest missile contract, does it follow that I believe that the actual world, or some mental model of mine, is constituted of such entities as the average man, foibles, Joe Sixpack, priorities, and inner tracks? When the press reports that a comet is aiming towards Jupiter and that lobster fishermen are overfishing New England waters, does that mean that the writers and readers think that comets have intentions and lobsters are fish?

These are questions of fact about the architecture of the mind, improperly formulated no doubt, because so little is understood.

If intuition is any guide, there seems to be a considerable gap between the semantic resources of language literally interpreted and thoughts expressed using them. I am happy to speak of the sun setting over the horizon, comets aiming directly at Jupiter, and waves hitting the shore, receding, and disappearing as the wind dies. But I’m not aware of having beliefs that correspond literally to the animistic and intentional terminology I freely use, or that conflict with anything I understand about relativity and the motions of molecules. Nor does the world, or my mental universe, seem to me to be populated by anything like what I describe as things that concern me. Psychologists and anthropologists exploring language-thought relations (e.g., the Sapir-Whorf hypothesis) find such problems hard and challenging; ready answers are offered in much of the contemporary philosophical literature, but on grounds that seem to me less than persuasive.

In fact, radically different answers are offered. Take language. Donald Davidson (1990) writes that “we all talk so freely about language, or languages, that we tend to forget that there are no such things in the world; there are only people and their various written and acoustical products. This point, obvious in itself, is nevertheless easy to forget ...”. To most philosophers of language, it is equally obvious that there *are* such things in the world as languages: indeed, “common, public languages”—Chinese, German, etc.—of which, some hold, we have “a partial, and partially erroneous, grasp” (Dummett 1986). Hilary Putnam (1989,

1993), among many others, takes the alleged fact to be as obvious as its denial is to Davidson, along with equally obvious facts about the things in the world that correspond to noun phrases rather freely, so it seems, so that the world contains whatever we might refer to as something that interests or bothers us, including the alleged denotata of words we do not know.¹⁵

A third position is that conclusions about such matters are rarely obvious: answers have to be found case by case, and the questions require more careful formulation in the first place. The ethnoscientist seeks to determine what people take to be constituents of the world, however they may talk about it. A different inquiry seeks the best theory of language and its use, and the states, processes, and structures that enter into it.

The questions arise in the simplest cases: nameable objects, substances, artifacts, actions, and so on. I take the thing in front of me to be a desk, but could be convinced that it is a hard bed for a dwarf that I am misusing as a desk; that's a matter of designer's intent and regular use. From one point of view, I take it to be the same thing whatever the answer, from another point of view, a different thing. Factors entering into such choices are diverse and complex. I take the contents of the cup on the desk to be tea, but if informed that it came from the tap after passing through a tea filter at the reservoir, I conclude that it is really water, not tea. Again, it is the same thing for me in either case from one point of view, a different thing from another. Some sticks I pass on the road are not a thing at all, unless it is explained to me that they were specifically constructed as some kind of object, whether by people or, perhaps, beavers. What is a thing, and if so what thing it is, depends on specific configurations of human interests, intentions, goals and actions, an observation as old as Aristotle. It could be that in such cases I do not change my beliefs about the constituents of the world as identification changes—that in my own variant of “folk science”, the entities that hold up my computer and fill the cup, and that I pass on the road, remain as they were independent of the explanations, which place them in unexpected relations to designs, intentions, uses, and purposes.

As the study of the language faculty and other cognitive systems progresses, we may come to understand in what respects my picture of the world is framed in terms of things selected and individuated by properties of my lexicon, or even involves entities and relationships describable at

¹⁵ That Putnam and Davidson differ is not entirely clear, since Putnam does not indicate what he means by “language” while Davidson spells out a notion modelled on formal language that is surely not Putnam's, though Davidson's conclusion would seem to exclude whatever is intended. Internalist linguistics would also be excluded unless we understand “people” to include their faculties, states, etc.

all by the resources of the language faculty. Some semantic properties do seem specifically linked to language, developing as part of it, closely integrated with its other aspects, even represented in natural ways within its morphological and syntactic structures. Terms of language may indicate positions in belief systems, which enrich further the complex perspectives they afford for viewing the world. Some terms, particularly those lacking internal relational structure, may do little more than that; notably “natural kind terms”, though the phrase is misleading, since they have little if anything to do with the kinds of nature. Akeel Bilgrami observes that analysis of lexical resources in terms of “a-linguistic agent’s perspective on things”, resisting dubious notions of independent reference, leads naturally to linking the study of meaning to “such things as beliefs as mediating the things in the world with which we stand in causal relations” and to the “radically local or contextual” notion of content that he develops in rejecting “the entire current way of thinking which bifurcates content into wide and narrow”. These seem to me fruitful directions to pursue.¹⁶

The study of semantic resources of the language faculty is not ethno-science, and both enterprises, of course, are to be distinguished from naturalistic inquiry into the range of topics that natural language and folk science address in their own ways. The observation is a truism in the case of falling apples, plants turning toward the light, and rockets aiming toward the heavens; here no one expects ordinary language or folk science to enter into attempts to gain theoretical understanding of the world, beyond their intuitive starting points. In contrast, it is considered a serious problem to determine whether “mentalistic talk and mental entities [will] eventually lose their place in our attempts to describe and explain the world” (Burge 1992). The belief that mentalistic talk and entities will lose their place is “eliminationism” or “eliminative materialism”, which Burge identifies as a major strand of the effort “to make philosophy scientific”—perhaps wrong, but an important thesis.

Why it is important is unclear. If we replace “mental” by “physical” in the thesis it loses its interest: “physicalistic talk and physical entities” have long ago “lost their place in our attempts to describe and explain the world”, if by “physicalistic” and “physical” we mean the notions of common discourse or folk science, and by “attempts to describe and explain the world” we mean naturalistic inquiry. Why should we expect anything different of “mentalistic talk and mental entities”? Why, for example, should we assume that psychology “seeks to refine, deepen, generalize and systematize some of the statements of informed common

¹⁶ Bilgrami, comments in Chomsky et al. (1993); Bilgrami (1992). On natural kind terms, see Bromberger (1992).

sense about people's mental activity" (Burge 1988),¹⁷ though chemistry, geology, and biology have no comparable concerns. No one expects ordinary talk about things happening in the "physical world" to have any particular relation to naturalistic theories; the terms belong to different intellectual universes. These facts are not taken to pose a body-body problem, nor has anyone proposed a thesis of "anomalism of the physical" to deal with them. The same should, then, be true of such statements as "John speaks Chinese" or "John took his umbrella because he expected rain"—though one may hope, in all cases, that science might yield some understanding and insight in the domains opened to inquiry by common sense perspectives.

There seems no basis here for any mind-body problem and no reason to question Davidson's thesis that there are no psychophysical laws that connect mental and physical events in an appropriate explanatory scheme; for similar reasons, there are no physico-physical laws relating ordinary talk about things to the natural sciences, even if the particular events described fall within their potential descriptive range. Distinctions between mental and other aspects of the world, in these respects, seem unwarranted, except in one respect: our theoretical understanding of language, mind, and people generally is so shallow, apart from limited domains, that we can only use our intuitive resources in thinking and talking about these matters.

It is not that ordinary discourse fails to talk about the world, or that the particulars it describes do not exist, or that the accounts are too imprecise. Rather, the categories used and principles invoked need not have even loose counterparts in naturalistic inquiry. That is true even of the parts of ordinary discourse that have a quasi-naturalistic cast. How people decide whether something is water or tea is of no concern to chemistry. It is no necessary task of biochemistry to decide 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 common sense notions would matter no more than for *the heavens*, or *energy*, or *solid*. Whether ordinary usage would consider viruses "alive" is of no interest to biologists, who will categorize as they choose in terms of genes and conditions under which they function. We cannot invoke ordinary usage to judge whether François Jacob is correct in telling us that "for the biologist, the living begins only with what was able to constitute a genetic program", though "for the chemist, in contrast, it is somewhat arbitrary to make a demarcation where there can only be continuity" (Jacob, 1973). Similarly, the concept "human being", with its curious properties of

¹⁷ Burge is describing what he takes to be "psychology as it is", but the context indicates that more is intended. On the assumption, see below p. 53f.

psychic continuity, does not enter the natural sciences. The theory of evolution and other parts of biology do try to understand John Smith and his place in nature; not, however, under the description “human being” or “person” as construed in ordinary language and thought. These notions are interesting for natural language semantics and ethnoscience, but not for the branches of human biology that seek to understand the nature of John Smith and his conspecifics or what distinguishes them from apes and plants.¹⁸

The special sciences too go their own ways. To borrow Jerry Fodor’s example of a meandering river eroding its banks, the earth sciences do not care under what circumstances people take it to be the same river if the flow is reversed or it is redirected on a different course, or when they regard something projecting from the sea as an island or a mountain with a watery base. The same should be expected in the case of such notions as *language* and *belief*, and terms of related semantic fields in various languages and cultural settings.

The particular natural sciences are commonly recognized to be largely artifacts and conveniences, which we do not expect to carve nature at its joints. The observation is uncontroversial for the “hard sciences”, but has been strongly challenged in the case of language. There has been much heated debate over what the subject matter of linguistics *really is*, and what categories of data are permitted to bear on it. A distinction is made between *linguistic evidence* that is appropriate for *linguistics*, versus *psychological* and other evidence that is not. Such discussions, which can be found in all the relevant disciplines, are foreign to naturalistic inquiry. An empirical observation does not come with a notice “I am for *X*”, written on its sleeve, where *X* is chemistry, linguistics, or whatever. No one asks whether the study of a complex molecule belongs to chemistry or biology, and no one should ask whether the study of linguistic expressions and their properties belongs to linguistics, psychology, or the brain sciences.

Nor can we know in advance what kinds of evidence might bear on these questions. Thus some current research suggests that studies of electrical activity of the brain may provide evidence bearing on them, a conceptual impossibility according to a considerable part of the literature, which also puts forth other odd contentions: for example, that studies of perceptual displacement of clicks might provide evidence about phrase boundaries, whereas observations about anaphora in Japanese, which provide far stronger evidence on naturalistic grounds, do not constitute evidence for factual theses at all because of some lethal form of indeterminacy (Quine 1992). Or that we should keep to—or even be interested in—“Grandma’s view” about the domain of linguistics, though presuma-

¹⁸ See Putnam (1993) for a contrary view with regard to these examples.

bly not chemistry (Devitt and Sterelny 1989). Or that studies of processing, acquisition, pathology, injury, genetic variability, and so on, cannot in principle be used as evidence about the existence and status of elements of linguistic representation (Scott Soames), contrary to what practising linguists have long believed; e.g., Edward Sapir and Roman Jakobson in classic work, or recent studies of priming effects in processing and their implications concerning unarticulated elements. All such moves reflect some form of dualism, an insistence that we must not treat the domain of the mental, or at least the linguistic, as we do other aspects of the world.

Methodological dualism has sometimes apparently been explicitly advocated. Consider Dummett's thesis that scientific accounts fall short of philosophical explanation for conceptual reasons. To take his example, suppose that a naturalistic approach to language succeeds beyond our wildest dreams. Suppose it provides a precise account of what happens when sound waves hit the ear and are processed, is fully integrated into a scientific theory of action, and solves the unification problem, integrating the theories of cells and computational processes. We would then have a successful theory of what Jones knows when he has acquired a language: what he knows about rhyme, entailment, usage appropriate to situations, and so on. But no matter how successful, Dummett writes, these discoveries would "contribute nothing to philosophy", which requires an answer to a different question: not how knowledge is stored or used, but "how it is delivered". The naturalistic account would be a "psychological hypothesis", but not a "philosophical explanation", because it does not tell us "the form in which [the body of knowledge] is delivered" (Dummett, 1993, p. xi; 1991 p. 97). For the sciences, the account tells everything that can be asked about the form in which knowledge is delivered, but philosophy calls for a kind of explanation unknown in naturalistic inquiry.

So understood, philosophy appears to exclude much of the core of traditional philosophy: Hume, for example, who was concerned with "the science of human nature", and sought to find "the secret springs and principles, by which the human mind is actuated in its operations", including those "parts of [our] knowledge" that are derived "by the original hand of nature", an enterprise he compared to Newton's. Had Hume achieved these goals, he would have established "psychological hypotheses", in Dummett's terms, but would not yet have contributed anything to philosophy. "Philosophical explanation" requires something more than a discovery of the "secret springs and principles" of the mind and how they function.

If I understand Dummett, philosophical explanation crucially involves access to consciousness. Imagine then a Martian creature *M* exactly like

us except that *M* can become aware of how its mind is “actuated in its operations”. When we ask *M* whether it is following the rules of phonology in constructing rhymes, or Condition (*B*) of Binding Theory in determining referential dependence, *M* reflects and says (truly), “Yes, that’s just what I’m doing”—by assumption, exactly what you and I are doing. For *M*, we would have a “philosophical explanation”; we would understand the form in which the knowledge is delivered, and could properly attribute knowledge to *M*. But we would not have crossed the bridge to “philosophical explanation” and attribution of knowledge for the human who operates exactly as *M* does, though without awareness. As Quine, John Searle, and others put it, we would be allowed to say that *M* is following rules and is guided by them, whereas the human cannot be described in these terms. To avoid immediate counterintuitive consequences, Searle insists further on a notion of “access in principle” that remains entirely obscure (see Chomsky forthcoming).

Are these proposals substantive or merely terminological? The latter, it seems to me; I do not see what substantive issue arises. It might be added that the proposals radically deviate from ordinary usage, for whatever that may be worth. In informal usage, we say that my granddaughter is following the rules for regular past tense and certain irregular verbs when she says “I rided my bike and brang it home”, though these rules are not accessible to consciousness, for children or adults, any more than those that Quine, Searle, and others disqualify. Saul Kripke’s “Wittgensteinian” concept of rule-following in terms of community norms is virtually the complement of ordinary usage, which typically attributes rule-guided behavior in cases of deviation from such norms, as in the example just given. In contrast, only a linguist would be likely to say that my granddaughter is following the rules of Binding Theory, conforming to the community (in fact, the human community, very likely).

In the study of other aspects of the world, we are satisfied with “best theory” arguments, and there is no privileged category of evidence that provides criteria for theoretical constructions. In the study of language and mind, naturalistic theory does not suffice: we must seek “philosophical explanations”, delimit inquiry in terms of some imposed criterion, require that theoretical posits be grounded in categories of evidence selected by the philosopher, and rely on notions such as “access in principle” that have no place in naturalistic inquiry.

Whatever all this means, there is a demand beyond naturalism, a form of dualism that remains to be explained and justified.

Philosophical demands are sometimes motivated by the problems of error and first-person authority. Defending a position much like the one advanced here, Barry Smith (1992, pp. 134-39) concludes that it still falls

short of “a philosophically satisfying account” for such reasons; it fails to “tell us what counts as using... words correctly, i.e., in accordance with certain normative patterns of use”, and to account for our authoritative knowledge of syntax and meaning in our own language. So “philosophical work... is vital to complete the overall project”, work that goes beyond “scientific psychology” (including internalist linguistics).

These conclusions seem to me unwarranted. Consider a typical example. Suppose that Peter, a normal speaker of English, says “John expects to like him”. I conclude that he intends to refer to two different people: John, and someone else picked out by the pronoun *him*. If Peter embeds the same expression in the context “Guess who—”, so that he said “Guess who John expects to like him”, I do not know whether or not he intended to refer only to John. In “John expects to like him”, *him* is not referentially dependent on *John*; in “Guess who John expects to like him”, the question is open. There is a good explanation of such facts in terms of an internalist linguistic theory, call it *T*.

Suppose *T* to be true of the Martian *M* and of us. *M* can tell us that he draws these conclusions on the basis of *T*, which he can recognize and even articulate; I cannot, though I operate exactly as *M* does. Given *M*’s conscious access to the rules it follows, some are inclined to feel that we have an account of *M*’s being “effortlessly authoritative” about the facts informally described; but the internalist naturalistic account “makes a puzzle” or a “total mystery” of this first-person authority in Peter’s case. Lacking *M*’s conscious access, Wright (1989) asks, how can Peter “understand... particular expressions”, say the ones in question, about which he is “effortlessly authoritative”.

Suppose that we put the matter differently. The kind of account that can be offered today, including *T*, does not “make a mystery” of first person authority, though it does *leave* a mystery, about both *M* and Peter. For both, we have an account that meets the conditions of the sciences (questions of precision and accuracy aside), but we lack any insight into the nature of consciousness, something not relevant to the matter of rule-following and first-person authority, though interesting in its own right.

Peter follows the rules of *T* because that is the way he is constructed, just as he sees the setting of the sun and the waves dashing against the rocks; his first-person authority is exhausted by this fact. As for what we call “error”, there are many possible kinds. Peter may depart from some external standard—say, using “disinterested” to mean “uninterested”, or using his native dialect in a formal lecture. He may choose to violate the rules, perhaps using the word “chair” to mean *table* in a code—knowing that in his own language it means *chair*. In doing so, he makes use of faculties of mind beyond the language faculty. He may misinterpret an

expression, in that his performance system yields an interpretation different from the one his internal language imposes; there are well-known categories of such cases, which have been fruitfully studied. Running through other possibilities, we seem to find no relevant limits to internalist psychology.

Others use different terms for what seem to be the same points. Thus Thomas Nagel argues that a full naturalistic theory of language, its use and acquisition, would not describe a “psychological mechanism” but “simply a physical mechanism—for it is incapable of giving rise to subjective conscious thought whose content consists of those rules themselves”. The crucial distinction, again, lies in access to consciousness in principle. The point seems the same as Dummett’s, but with different terminology: “psychological” replacing “philosophical”. Here the problem of understanding “access in principle” and “content of thought” is compounded by the obscurity of the notion “physical mechanism”, which had some meaning in pre-Newtonian physics, but not since (see Nagel 1993).

Unless offered some new notion of “body” or “material” or “physical”, we have no concept of naturalism apart from methodological naturalism. More conventional usage refers to a different doctrine: “metaphysical naturalism”, which Burge in his historical review describes as “one of the few orthodoxies in American philosophy” in recent years; in other variants materialism, physicalism, eliminativism, “the naturalization of philosophy”, and so on. These doctrines are intelligible only insofar as the domain of the physical is somehow specified.

One leading advocate, Daniel Dennett, formulates the doctrine in this way: the “naturalization of philosophy”, which he describes as “one of the happiest trends in philosophy since the 1960s”, holds that “philosophical accounts of our minds, our knowledge, and our language must in the end be continuous with, or harmonious with, the natural sciences”. In a discussion of contemporary naturalism, T.R. Baldwin (1993) cites this statement to illustrate the thesis of “metaphysical naturalism”. Like other formulations, it poses some problems. What are “philosophical accounts” as distinct from others, particularly in this “naturalized” sense of philosophy? And what are the natural sciences? Surely not what is understood today, which may not be “continuous and harmonious” with tomorrow’s physics. Some Peircean ideal, perhaps? That doesn’t seem promising. What the human mind can attain in the limit? That at least is a potential topic of inquiry, but it leaves us in even worse shape in the present context. If “metaphysical naturalism” is understood as a hope for eventual unification of the study of the mental with other parts of science, no one could disagree, but it is a thesis of little interest, not “a happy trend in philosophy”.

Take the version of this doctrine expressed by Quine, whom Burge identifies as the source of the contemporary orthodoxy. In his most recent formulation, the “naturalistic thesis” is that “the world is as natural science says it is, insofar as natural science is right”. What is “natural science”? Quine’s total answer is: “theories of quarks and the like”. What counts as like enough? There are hints at answers but they seem completely arbitrary, at least by ordinary naturalistic criteria (see Quine 1992, and Chomsky forthcoming).

Suppose we identify the mind-body problem (or perhaps its core) as the problem of explaining how consciousness relates to neural structures. If so, it seems much like others that have arisen through the history of science, sometimes with no solution: the problem of explaining terrestrial and planetary motion in terms of the “mechanical philosophy” and its contact mechanics, demonstrated to be irresolvable by Newton, and overcome by introducing what were understood to be “immaterial” forces; the problem of reducing electricity and magnetism to mechanics, unsolvable and overcome by the even stranger assumption that fields are real physical things; the problem of reducing chemistry to the world of hard particles in motion, energy, and electromagnetic waves, only overcome with the introduction of even weirder hypotheses about the nature of the physical world. In each of these cases, unification was achieved and the problem resolved not by reduction, but by quite different forms of accommodation. Even the reduction of biology to biochemistry is a bit of an illusion, since it came only a few years after the unification of chemistry and a radically new physics.

These examples do differ from the consciousness-brain problem in one important way: it was possible to construct intelligible theories of the irreducible phenomena that were far from superficial, while in the case of consciousness, we do not seem to progress much beyond description and illustration of phenomena (Freudians, Jungians, and others might disagree). The matter is seen more sharply in the case of language. The normal use of language involves a “creative aspect” which, for the Cartesians, provided the best evidence for the existence of other minds. Neither the computational properties of the language faculty nor the creative aspects of use can be related in interesting ways to anything known about cells, but the two topics differ in that for the computational properties, there are intelligible explanatory theories, while for the creative aspects of use, there is only description and illustration. If so, the crucial issue is not real or apparent irreducibility, a common phenomenon in the history of science, but the fact that we can only stare in puzzlement at such aspects of mind as consciousness and expression of thought that is coherent and

appropriate but uncaused, a characteristic feature of core problems of philosophy, as Colin McGinn (1993) has argued.

Furthermore, apart from the fact that literal reduction is hardly the norm as science has proceeded towards unification, there is uncertainty as to whether it even makes sense as a project. Silvan Schweber (1993) writes that recent work in condensed matter physics, which has created phenomena such as superconductivity that are “genuine novelties in the universe”, has also raised earlier skepticism about the possibility of reduction to “an almost rigorously proved assertion”, leading to a conception of “emergent laws” in a new sense. Whatever the validity of the conclusion, it is at least clear that philosophical doctrines have nothing to say about it; even less so in the domain of mind and brain, where vastly less is understood.

A naturalistic approach simply follows the post-Newtonian course, recognizing that we can do no more than seek the best theoretical account of the phenomena of experience and experiment, wherever the quest leads.

As in other branches of science, we expect to leave the concepts of common sense understanding behind. Take a concrete example, the case of a woman called “Laura” studied by Jeni Yamada (1990). Laura’s language capacities are apparently intact, but her cognitive and pragmatic competence is limited. She has a large vocabulary that she uses in appropriate ways, though apparently without much understanding. Yamada suggests the analogy of young children who use colour words in the proper places “to dress up discourse”, but without grasping their referential properties. Laura knows when she should describe herself and others as sad or happy, but apparently without capacity to feel sad or happy; she’s a kind of behaviourist. Does she *know* or *understand* or *speak* English? The question is meaningless. Usual assumptions about people do not hold in Laura’s case; the presuppositions of ordinary usage are not satisfied. Naturalistic theories of language and mind may provide concepts that apply to Laura, but these depart from ordinary language. These concepts, incidentally, are part of an internalist theory of language and mind, the only kind we have. We cannot ask, for example, about the “broad content” of Laura’s speech unless the technical notion is extended to this case.

Take a somewhat different case: my four-year-old granddaughter. Does she speak English? What we say in ordinary discourse is that she has a partial knowledge of the language that she will ultimately attain if events follow the expected course, though what she now speaks is not a language at all. But if all adults were to die, and children her age were miraculously to survive, what they speak would be perfectly normal human languages, ones not found today. This teleological aspect of the common sense notion of language is among the many curious and complex features that render

the concept inappropriate for the attempt to understand language and its use, just as biology does not concern itself with the psychic continuity of persons and the earth sciences do not care what people call the same river, or a mountain or an island. These are truisms in the case of “the physical”; likewise for “the mental”, dualistic assumptions aside.

The same holds of attribution of belief. It is a reasonable project of natural science to determine whether people (in particular, young children) interpret what happens in the world in terms of such notions as belief and desire, falling from the heavens toward the earth, turning toward the light, and so on; and the conditions under which they use such intentional and objectual discourse in various languages (perhaps a different matter, as noted). Quite independently, we may ask whether the theory of people, meteors, and flowers should involve such notions. The current answer is “definitely not” in the case of flowers and meteors, and unknown in the case of people, because we do not know much at all. Let us consider a third kind of problem, which does not fall within either framework: the problem of determining when we *should* attribute belief, or rising and turning and aiming toward—when we *are justified* in doing so? To quote one recent formulation, we ask what are “the philosophically necessary condition[s] of being a true believer”. Access to consciousness is usually invoked at this point, and Quinean indeterminacy is commonly held to arise for belief, though not the other cases, for which no “philosophical demand” is raised at all (Clark and Karmiloff-Smith 1993). No one seeks to clarify the philosophically necessary conditions for a comet to be truly aiming at the earth—failing to hit it, if we are lucky (another intentional attribution).

Similarly, we are invited to explore the criteria for determining where to draw the line between comets aiming at the earth and Jones walking toward his desk; on which side should we place barnacles attaching to shells and bugs flying toward the light? Such questions do not belong to ethnoscience or the study of the lexicon, nor to naturalistic inquiry in other parts of the sciences. Again, it seems that the quest is for “philosophical explanations”, whatever they may be.

The same questions arise about debates over manifestation of “intelligence” and “language use”. In the case of vision, locomotion, and other systems one might seek homologies or evolutionary connections. But mental properties are not approached in such ways. Something different is at stake in the debates about whether machines think, or translate Chinese, or play chess. We ask whether an imagined Martian or a programmed computer could understand Chinese, but not whether an extra-terrestrial creature or a camera could see, like humans. There is a substantial literature on whether a person mechanically carrying out an

algorithm with coded inputs and outputs can properly be said to be translating English to Chinese, but none on the analogous questions that could be raised about mimicking the computations and algorithms that map retinal stimulation to visual image or reaching for an object. It is taken to be a crucial task for the *theory of meaning* to construct notions that would apply to any creature however constituted, real or imagined; but this is not a task at all for the theory of vision or locomotion. Curiously, this is also not considered a task for the theory of phonology, though the questions have as much merit here—none, I think. Similarly, no one asks what would count as a circulatory system, or a molecule, in some world of different objects or different laws of nature.

The discussions are not only dualistic in essence, but also, it seems, without any clear purpose or point: on a par with debates about whether the space shuttle flies or submarines set sail, but do not swim; questions of decision, not fact, in these cases, though assumed to be substantive in the case of the mind, on assumptions that have yet to be explained—and that, incidentally, ignore an explicit warning by Turing (1950) in the classic paper that inspired much of the vigorous debate of the past years.

When we turn to language, the internalism-externalism issues arise; though again only for the theory of meaning, not for phonology, where they could be posed in the same ways. Thus we are asked to consider whether meanings are “in the head”, or are externally determined. The conventional answer today is that they are externally determined by two kinds of factors: features of the real world, and norms of communities.

What notion of meaning is being investigated? Rational reconstruction of actual translation practice is a goal sometimes suggested, but proposals are not seriously evaluated in these terms and the significance of the project is also unclear. Another stated goal is to determine the meaning of a word (but apparently, not the sound of a word) in a “shared public language”, a notion that remains to be formulated in some coherent terms.¹⁹ Plainly, the goal is not to discover the semantic features of the word “meaning” in English or similar expressions, if they can be found, in other languages. Does the inquiry belong to ethnoscience, an investigation of our conceptual resources? The inquiries that are conducted do not seem well-designed for this purpose. The questions also do not have to do with naturalistic inquiry into the nature of language and its use, which will develop in its own ways.

What other possibility is there? The answer is not clear.

In fact, some curious moves take place at this point. Consider the Twin-Earth thought experiment designed by Hilary Putnam, which has provided much of the motivation for externalist assumptions. In one

¹⁹ These motives lie behind Putnam (1975), as he reiterates in Putnam (1993).

version, we are to explore our intuitions about the *extension* or *reference* of the word “water” on Twin-Earth, where speakers identical to us use it to refer to XYZ, which is not H_2O . But we can have no intuitions about the question, because the terms *extension*, *reference*, *true of*, *denote*, and others related to them are technical innovations, which mean exactly what their inventors tell us they mean: it would make as little sense to explore our intuitions about tensors or undecidability, in the technical sense.

Suppose we pose the thought experiment using ordinary language. Suppose, for example, that Twin-Oscar comes to earth, is thirsty, and asks for *that*, pointing either to a glass of Sprite or of what comes from the faucet—some odd mixture of H_2O , chlorine, and I hate to think what else, differing significantly from place to place (but called “water”). Is he making a mistake in both cases? In one case? Which one? Suppose he refers to stuff from the faucet that passed through a tea filter at the reservoir (and therefore is *water* for Oscar), and to the chemically identical substance that had a tea bag dipped into it (so it is not *water* for Oscar, but *tea*). In which case (if either) is Twin-Oscar mistaken? Turning to “content of belief”, if Twin-Oscar continues to ask for what comes from the faucet to quench his thirst, calling it “water”, has he changed his beliefs about water—irrationally, since he has no evidence for such a change? Or is he behaving rationally, keeping his original beliefs about water, which allow for the stuff on Earth to be water (in Twin-English) in the first place? If the latter, then beliefs about water are shared on Earth and Twin-Earth, just as on either planet, beliefs may differ about the very same substance, taken to be either water or tea as circumstances vary, even with full and precise knowledge that the objects of the different beliefs have exactly the same constitution.

I have my intuitions, which would be relevant to the study of the lexicon and ethnoscience, but which undermine the intended conclusions of the thought experiment.

There are numerous other problems. The Twin-Earth problem is posed by withdrawing the presuppositions of discourse on which normal usage rests. It is akin to asking whether Laura understands English. Furthermore, if the argument applies to “water”, then why not to “earth”, “air”, and “fire”, which had a comparable status in one early tradition? What is “same substance” in these cases? Or consider “the heavens”. I use the term with an indexical character, to refer to what I see on a cloudless night: something different in Boston and Tasmania. With ordinary presuppositions withdrawn, as on Twin-Earth, I might decide (in some circumstances) to use “water” the same way. The dimensions of choice are so varied that it is not surprising that “most ears not previously contaminated by philosophical theory” provide no clear judgments in the standard

cases, as Stich (1993) has observed. That would not be a decisive objection in a richer theoretical context, but it is a warning sign that should not be ignored when we have little beyond alleged examples.

Putnam's response to such problems seems to me unconvincing (Putnam 1993). He agrees that words do not refer, so intuitions about *reference of words* have to be reformulated in some different way. He adopts the Peircean position that "*reference* [in the sense of 'true of'] is a triadic relation (person *X* refers to object *Y* by sign *S*)", where the *Y*s are "real objects in the world". Furthermore, "That there is a relation between our words and things in the world is fundamental to our existence; thought without a relation to things in the world is empty."²⁰ Thus a word refers to (is true of) a real object in the world when people use the word to refer. Since people use the word "Chinese" to refer to the language spoken in Beijing and Hong Kong, that is "a real object in the world", and the same should apparently hold of "the mind", "the average man", "Joe Sixpack", "free trade", "the heavens", etc., as well as of adjectives, verbs, and other relational expressions.

Such super-Whorfian conclusions aside, several problems arise. First, accepting this formulation, the externalist arguments collapse, including the Twin-Earth experiment, the case of "the division of linguistic labor",²¹ and others. When Twin-Oscar, visiting Earth, asks for a cup of water, referring to what is in the cup as "water", then we conclude, following Putnam's revision, that *water* in Twin-English is true of H₂O, so that meanings are back in the head. The other arguments fail for similar reasons.

Second, the revision is not helpful, since the Peircean thesis involves an invented technical notion of *reference*, so we are back where we were, with intuitions that we cannot have. In ordinary usage, "reference" is not a triadic relation of the Peircean sort. Rather, person *X* refers to *Y* by expression *E* under circumstances *C*, so the relation is at least tetradic; and *Y* need not be a real object in the world or regarded that way by *X*. More generally, person *X* uses expression *E* with its intrinsic semantic properties to talk about the world from certain intricate perspectives, focusing attention on particular aspects of it, under circumstances *C*, with the "locality of content" they induce (in Bilgrami's sense). Indeed the components of *E* may have no intrinsic semantic relation at all to what Jones is

²⁰ I omit a footnote in which Putnam qualifies his claim. I believe that his statement about emptiness of thought seems much too strong, but put that aside.

²¹ A questionable term, since Putnam seems to have dropped the implicit requirement that the "experts" to whom we defer even speak our language; the social aspect therefore disappears, and we are back to "same substance" considerations.

referring to, as when he says the performance at Jordan Hall was remarkable, referring to Boston and his favourite string quartet.

Putnam writes that he thinks “Chomsky knows perfectly well that there is a relation between speakers, words, and things in the world”. So there sometimes is, abstracting from circumstances of use, in more or less the sense in which a relation holds of people, hands, and rocks, in that I can use my hand to pick up a rock. But that leaves us a long way from establishing anything remotely like the conclusions Putnam wants to reach.

From the natural language and common-sense concepts of *reference* and the like, we can extract no relevant “relation between our words and things in the world”. And when we begin to fill out the picture to approach actual usage and thought, the externalist conclusions are not sustained, except that in the welter of uses, some will have the desired properties; in special circumstances, we may indeed understand *water* in the sense of “same liquid”, where “liquid” and “same” are the kinds of notions that science seeks to discover, and satisfy other externalist assumptions. Thinking about the world is no doubt “fundamental to our existence”, but this does not seem to be a good way to gain a better understanding of the matter.

The philosophical inquiry seems oddly framed in other respects as well. Thus the word “water” is a collection of phonetic, semantic, and formal properties, which are accessed by various performance systems for articulation, perception, talking about the world, and so on. If we deny that its meaning is in the head, why not also that its phonetic aspects are in the head? Why does no one propose that the *phonetic content* of “water” is determined by certain motions of molecules or conventions about “proper pronunciation”? The questions are understood to be absurd or irrelevant. Why not also in the case of meaning?

The literature suggests some answers. Thus, Putnam’s conclusions about “water” and H_2O are in part motivated by the problem of intelligibility in scientific discourse. As he points out, we do not want to say that Bohr was talking utter nonsense when he used the term “electron” in pre-quantum theoretic days, or that all his statements were false. To avoid such absurd conclusions, Putnam argues that Bohr was referring to *real* atoms and electrons, which perhaps some experts can finally tell us about (or maybe not). If reference is determined by meaning, then meanings aren’t in the head, as Twin-Earth experiments are supposed to show.

The argument, however, is not persuasive, for reasons beyond those already mentioned. Jay Atlas (1989) has pointed out that nuclear engineers distinguish “light water” from “heavy water”, only the former being H_2O . Taking them as experts, have we been misusing “water” all along, really meaning light water? Pre-Avogadro, chemists were using

“atom” and “molecule” interchangeably. To render what they were saying intelligible, do we have to assume that they were referring to what are now called “atoms” and “molecules” (or what they *really* are, which no one today may know)? After the Bohr model of the atom was available, it was proposed that acids and bases be understood as potential acceptors or donors of electrons, which made boron and aluminium chlorides acids alongside of sulphuric acid, opening up “a whole new area of physical inorganic chemistry” (Brock 1992, p. 482). Were earlier scientists *really* referring to boron as an acid? Must we assume that to render their views intelligible?

To take a simpler example, closer to home, must we assume that structural phonologists, 40 years ago, were referring to what generative phonologists call phonological units, though they hotly denied it and rightly so? Structuralist phonology is surely intelligible; without assuming that there are entities of the kind it postulated, much of the theory can be reinterpreted today, with many results carried over.

What is required in all such cases is some degree of shared structure. In none of them is there any principled way to determine how much must be shared, or what “similarity of belief” is required. 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 theoretical understanding.

Putnam objects that mere structural similarity “is very different from saying that either theory *describes*, however imperfectly, the behavior of the elusive extra-mental phenomena we refer to as *electrons*”—or *light water*, *atoms* and *molecules*, *acids* and *bases*, *phonemes*, etc. That is true, but not relevant. In all cases, including the current theories, we have to add whatever it is that distinguishes theories about the world from science fiction. We take such theories to describe extra-mental phenomena, however imperfectly, whether they involve Apollo and the sun, Galen’s four humours and the atoms of Democritus, Descartes’s tubes with animal spirits, ... , and on to today’s attempts. But in no case is there any convincing reason to adopt a theory of *real reference* of the kind that has been based on externalist arguments of this nature.

These considerations aside, discussions about *reference* in the sciences have no particular bearing on human language and common sense understanding unless we add the further assumption that such words as “electron”, “base”, “eigenvector”, “phoneme”, and so on, belong to English and other natural languages, presumably along with expressions in which they appear, perhaps also formulas, diagrams, etc. Putnam has assumed that the lexicon is homogeneous in this sense. Thus in defending meaning

holism, he argues that the theory of meaning must deal with “the hardest case”; he gives the example of “momentum”, which was once defined in a way now taken to express a falsehood. However we interpret this, it has no bearing on the inquiry into language unless we assume that “momentum” in the physicist’s sense enters the lexicon by the same mechanisms of the language faculty that allow a child to pick up such words as “house” and “rise”, and has the properties of lexical entries determined by the language faculty. That seems dubious, to say the least.

Putnam (1993, p. 383) is right to say that I “agree that there is such a relation as reference”, in the technical sense, or at least may be, but misses my point: it is reasonable to suppose that naturalistic inquiry aims to construct symbolic systems in which certain expressions are intended to pick out things in the world.²² But there is no reason to believe that such endeavours inform us about ordinary language and common sense understanding. It seems to me surprising that Putnam should take the position he does, given his eloquent critique of “scientism”.

Putting meaning aside, are the contents of thought externally determined? We cannot sensibly ask such questions about *content*, wide or narrow, technical notions again. But we can ask whether we attribute thoughts to people on grounds that do not keep to their internal state. That we do is clear without exotic examples. If Jones tells me he is mourning those who died in the trenches at Verdun 50 years ago, I can properly say that he is really talking about (thinking of) World War I, not World War II; or, alternatively, that he is mistaken about World War II, which is what he is talking about (thinking of). In the first case, I am attributing to him a state that is not internal; the attribution is based on my beliefs, not his. There is no real question as to whether psychology deals with Jones’s state as specified in this case; that is again a question of decision, in this case, about the invented technical term “psychology”. Similarly, if Anna Karenina is modelled on a real person, Tolstoy might have been thinking, talking, having beliefs, etc., about her, and some of his knowledgeable readers as well; and as for Smith, who knows nothing about this, I might decide one way or another, as circumstances vary. However this turns out, it tells us nothing about the “real” subject matter of psychology, though these could be reasonable topics for internalist inquiry into how people talk about the world, inquiry that seeks to find out about the internal states that lead people to describe others in various ways as they interpret circumstances variously.

²² Irrelevantly here, it could be that a technical notion of *reference* should be introduced in the study of the syntax of mental representations, much as relations among phonetic features are introduced into phonology.

In this context too, the thought experiments designed to support anti-internalist conclusions often seem based on questionable assumptions. Take for example Lynne Rudder Baker's locust-cricket example, slightly simplified (Baker 1988). Suppose that Jones speaks ordinary English, and Smith does too except that in his speech community, crickets are called *locusts*. Suppose *J* learns his language from Jones, and *S* from Smith, and they learn the term "locust" from the same pictures, ambiguous between locusts and crickets, along with "information which by chance pertains to both locusts and crickets". Since the intentions of the instructors are different, it "seems straightforward", Baker concludes, that *J* has "acquired the belief that locusts are a menace and [*S*] acquired the belief that crickets are a menace", though *J* and *S* are in the same internal state

Under these assumptions, *J* and *S* will generalize the same way, so if presented with an unambiguous locust they will each call it "a locust", though *S* will be making an error because the beliefs he expresses are about crickets, not locusts. Suppose *S* moves to an island with speakers of an unrelated language, and his descendants learn exactly his language, indefinitely, all records and cognates having disappeared; similarly *J*. The *J* and *S* progeny are now indistinguishable in their language and its use, and the history is unrecoverable so they could never learn otherwise. Nevertheless, it should seem straightforward that they have different beliefs, and that the *S* progeny are making many errors in using their word "locust", always talking about and thinking of crickets. It could be, in fact, that we are of the *S*-progeny type, that somewhere in the mists of prehistory our ancestors acquired the word that became "locust" under the conditions of *S*, their instructor having intended to refer to some different species *X*, so that the beliefs we express using the word "locust" are really about *X*'s and are often mistaken.

Nothing of the sort seems at all straightforward to me, even the first step. But it's also not clear why it matters. Suppose we accept Baker's intuitions. What would this tell us about language, belief, and thought? At most, that sometimes we might attribute beliefs etc. to *X* in terms of other people's beliefs and intentions; but that is clear from simple and ordinary cases. Again, inquiry into the ways we attribute belief as circumstances vary is a legitimate topic of linguistic semantics and ethno-science, but the study of how people attain cognitive states, interact, and so on, will proceed along its separate course.

A standard externalist argument is that unless the external world determines the contents of the thought of an agent, "it is an utter mystery how that agent's thoughts can be publicly available to another" (Bilgrami 1992, p. 4). For psychology, the assumption is not needed. To account for the way Smith understands what Jones says we need not appeal to entities

in the external world that correspond to the phonetic representations in the mind of Smith and Jones (say, some kind of motions of molecules associated with the syntactic entity “bilabial stop”); and external objects are no more required in the case of meanings and thoughts. Other possibilities are certainly available, and probably correct. Thus it could be that Smith assumes that Jones is identical to him, modulo some modifications *M*, and then seeks to work out *M*, a task that may be easy, hard, or impossible. Insofar as Smith succeeds, he attributes to Jones the expression that his own mind constructs, including its sound and meaning, communication being a more-or-less affair.²³ And using a variety of other information, he seeks to ascertain Jones’s thoughts, perhaps in a similar way.

To be sure, this is psychology, and the issues are supposed to arise only in folk psychology, for Bilgrami at least. But the conclusions seem no better founded here. We have no reason to believe that Mary interprets the interactions of Smith and Jones by postulating “publicly available” entities that fix thoughts, meanings, or sounds. Furthermore, it is not clear that a mystery about communication would even be relevant to folk psychology, which need not and commonly does not face the task of resolving such problems.

Examples of the Twin-Earth type serve as one prong of conventional externalist theories of language and thought. The other prong involves deference to authority and experts, community norms, and so on. Meanings are said not to be “in the head” because they are fixed in such terms. Again, we may ask where the concept of meaning under investigation belongs. It is plainly not part of some scientific inquiry into language and its use, or into the lexical entry for “meaning” and “language” in English. Is it speculative ethnoscience, a study of “the commonsense psychological explanation of human behavior”, as Bilgrami (1992, p. 15) describes the project, while rejecting this prong of the argument (rightly, I believe)? Perhaps that is what is intended, but if so, the conclusions seem highly variable, as conditions vary, with nothing of much clarity emerging.

Whatever the inquiry may be about, it crucially relies on a notion of “common, public language” that remains mysterious. If it is the notion of ordinary discourse, it is useless for any form of theoretical explanation. In the empirical study of language, it has long been taken for granted that there is nothing in the world selected by such terms as “Chinese”, or

²³ It does not follow, however, that “meaning alike for us merely means, if anything, that we are communicating successfully” (Quine, unpublished manuscript, cited by Dreben 1993). Similarly, sounding alike for us does not merely mean that we are communicating successfully. In both cases, there is a good deal more to say about what is “alike” in terms of shared properties of language and mind, when we depart from Quine’s anti-naturalist behaviourist strictures.

“German”, or even much narrower ones. Speaking the same language is much like “living near” or “looking like”; there are no categories to be fixed. The fact that ordinary language provides no way to refer to what my granddaughter is speaking is fine for ordinary life, but empirical inquiry requires a different concept. In that inquiry, her language faculty is in a certain state, which determines (or perhaps *is*) her “language”. Communities, cultures, patterns of deference, and so on, are established in human life in all sorts of ways, with no particular relation to anything we call “languages” in informal discourse. There is no meaningful answer to the question whether Bert should refer to the pain in his thigh as arthritis; or whether he should use the word “disinterested” to mean “unbiased”, as the dictionary says, or “uninterested”, as virtually every speaker believes; or whether he should pronounce words as in Boston or London.²⁴

There is simply no way of making sense of this prong of the externalist theory of meaning and language, as far as I can see, or of any of the work in theory of meaning and philosophy of language that relies on such notions, a statement that is intended to cut a rather wide swath.

In brief, though naturalism does not entail an internalist approach, it does seem to leave no realistic alternative. In actual empirical inquiry, that approach is regularly adopted, even when that is denied, a matter I have discussed elsewhere; as is familiar, to determine what scientists are doing we investigate their practice, not what they say about it.

As noted earlier, the issue of legitimacy of inquiries that go beyond internalist limits does not arise. This should be the merest truism. Accordingly, I am constantly surprised to read that I and others deny it. Thus a recent text on sociolinguistics opens with the remarkable claim that “modern linguistics has generally taken for granted that grammars are unrelated to the social lives of their speakers”, an absurd idea, advocated by no one, which the author attributes to my insistence that “questions of power ... are not the sorts of issues which linguists should address” (Romaine 1994)—that I should not engage in activities that occupy a good part of my time and energy. The book ends with the conclusion that

²⁴ These observations, familiar in the study of language, should be distinguished from Davidson’s conclusion (1986) that “there is no such thing as a language” in the sense generally assumed by “philosophers and linguists”, “no such thing to be learned, mastered, or born with”. But Davidson has a very different notion of *language* in mind; and though he is surely right in thinking that “there is no such thing”, the argument for that conclusion or about the notions of the empirical study of language is flawed. He observes correctly that in actual communication, all sorts of conjectures are used in a “passing theory”, which is a psychological particular. But it does not follow that there is no use for “the concept of a language”, for a “portable interpreting machine set to grind out the meaning of an arbitrary utterance”, etc. That would be like arguing that there is no jet stream, because of the chaotic elements in weather patterns. See note 15; and Chomsky (1993a) for further comment.

“linguistic differences enact and transmit inequalities in power and status”—there are, for example, prestige dialects—a discovery that is held to refute my contention that the study of such matters is not illuminated by what is presently understood about the nature of language.

Similar pronouncements abound in the literature, often put forth with much passion and indignation. They appear to be based on a belief that I have indeed expressed: that people should tell the truth. In particular, they should not claim special insight in areas of human concern unless the claims are true; and if they are, they should impart that special knowledge, which is rarely difficult. Posturing about such matters merely serves to intimidate and marginalize, reinforcing “inequalities in power and status”. Furthermore, to make very clear the limits of understanding is a serious responsibility in a culture in which alleged expertise is given often unwarranted prestige. If inquiry in areas of basic human concern can draw from authentic discoveries about language, vision, or whatever, well and good, but that has to be shown, not proclaimed. As for sociolinguistics, it is a perfectly legitimate inquiry, externalist by definition. It borrows from internalist inquiry into humans, but suggests no alternative to it. How much its findings illuminate issues of power and status is a separate question.

To cite another case, Putnam (1993) interprets my comments (actually, truisms) about “shared public language” as implying that unless “cultures can be defined essentialistically”, we should “forget about them and return to the serious business of computer modelling”—by which he seems to mean naturalistic inquiry into the language faculty, to which computer modelling might make some contribution, though it has never been a particular interest of mine. But the problems faced by uncritical reliance on this notion are not overcome by invocation of “culture” or “cultural artifacts”; and recognition of simple facts about Chinese, English, etc., and about the irrelevance of culture to the matters in question, in no way suggests the conclusion he draws. Cultures cross-cut anything that might reasonably be called “languages” in all sorts of ways, and “cultural studies” leave the problems where they were.

Putnam’s statement that “*Languages and meanings are cultural realities*” (his emphasis) is accurate in one sense, which is why (like everyone else) I describe the way the terms are understood in the cultures we more or less share in terms of structures of power and authority, deference patterns, literary monuments, flags and (often mythical) histories, and so on. Such terms as “language” are used in different ways in other speech communities; and our terms *belief*, *meaning*, etc., commonly lack any close counterpart. But these “cultural realities” do not contribute to understanding how language is acquired, understood, and used, how it is con-

stituted and changes over time, how it is related to other faculties of mind and to human action generally. Neither the empirical study of language itself, nor Putnam's "cultural studies (history, anthropology, sociology, parts of philosophy)", when seriously pursued, make use of the notion of "shared public language" of ordinary usage, apart from informal comment; in various contexts, an anthropologist may speak of the Chinese, or Chinese-Japanese, or East Asian culture area, of the culture of scientists speaking entirely different languages, of the culture of slum-dwellers in New York, Cairo, and Rio, and so on, in an intricate array that lacks any interesting relation to the languages spoken, or what are called "languages" in ordinary usage or in our literary cultures and others.

Such languages often are "cultural artifacts" in a narrower sense: partially invented "standard languages" that few may speak and that may even violate the principles of language. It is in terms of such artifacts that "norms" and "correct usage" are determined in many cultures, matters of little interest to "cultural studies", if only because they are too transparent. There is little interest in studying the behavior of the French Academy, for example.

In cultural studies, as in informal usage, we say, perfectly intelligibly, that John speaks the same language as Bill, looks like Bill, and lives near Bill. But we are not therefore misled into believing that the world is divided into objective areas or places, or that there is a shape that John and Bill share; or a common language. The problem is not open texture or lack of "sharp boundaries", as Putnam believes, any more than in the case of "area" or "era". "Standard languages" are in fact quite sharply determined (e.g., by the French Academy). In other usages too the boundaries of "language" are reasonably sharp, as these things go, determined by such matters as colors on maps and the like. But ordinary usage provides no notion of "shared public language" that comes even close to meeting the requirements of empirical inquiry or serious philosophical reflection on language and its use, and no more adequate notion has been proposed. Nor is there an explanatory gap that would be filled by inventing such a notion, as far as is known.

A central point of the article on which Putnam is commenting is that "Many questions, including those of greatest human significance one might argue, do not fall within naturalistic inquiry; we approach them in other ways". There is no implication there, or elsewhere, that we should keep to "the serious business of computer modelling", but only that we should keep to "serious business", whatever the domain.

Is there a problem with internalist (or individualist) approaches to other domains of psychology? So it is widely claimed, but on dubious grounds, I think. Take the study of hearing. One long-standing question is how the

auditory cortex determines the location of a sound. There does not seem to be any “auditory map”, as there is a visual and somatosensory map. Some recent work suggests that the auditory cortex registers sound location not by spatial arrangement of neurons, but by a temporal pattern of firing in a kind of “Morse code” (Barinaga 1994, Middlebrooks, et al. 1994). The discussion is worded in the usual mixture of technical and informal discourse. Someone reading it might be misled into thinking that the theory of auditory perception is externalist, making crucial reference to “solving problems” posed by the external world of sounds. But that is an illusion. The auditory system doesn’t “solve problems” in any technical sense of this term, and if they knew how to do so, the researchers might choose to stimulate the receptors directly instead of using loud-speakers—much as they did in the computer model which, in fact, provided the main evidence for their theory of sound localization, which would work as well for a brain in a vat as for an owl turning its head to face a mouse in the brush.

The same considerations apply to the study of visual perception along lines pioneered by David Marr, which has been much discussed in this connection. This work is mostly concerned with operations carried out by the retina; loosely put, the mapping of retinal images to the visual cortex. Marr’s famous three levels of analysis—computational, algorithmic, and implementation—have to do with ways of construing such mappings. Again, the theory applies to a brain in a vat exactly as it does to a person seeing an object in motion. The latter case has indeed been studied, in work of Marr’s collaborator Shimon Ullman. His studies of determination of structure from motion used tachistoscopic presentations that caused the subject to see a rotating cube, though there was no such thing in the environment; “see”, here, is used in its normal sense, not as an achievement verb. If Ullman could have stimulated the retina directly, he would have done that; or the optic nerve. The investigation, Ullman writes, “concerns the nature of the internal representations used by the visual system and the processes by which they are derived” (Ullman 1979, p. 3). The account is completely internalist. There is no meaningful question about the “content” of the internal representations of a person seeing a cube under the conditions of the experiments, or if the retina is stimulated by a rotating cube, or by a video of a rotating cube; or about the content of a frog’s “representation of” a fly or of a moving dot in the standard experimental studies of frog vision. No notion like “content”, or “representation of”, figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as “a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it consists of

arrays of image intensity values as detected by the photoreceptors in the retina" (Marr 1982, p. 31)—where "representation" is not to be understood relationally, as "representation of".

Technical presentations talk about algorithms "breaking down" under some conditions, and giving the "correct answer" in others, where the "correct answer" may be, for example, the strong three-dimensional percept given by a random dot stereogram. They may also speak of "mis-perception" in the case of the person or frog in the experiments, though perhaps not when a photoreceptor on a street light is activated by a search-light rather than the sun. And they speak of the brain as "solving problems" and as "adapted to normal situations" in which the visual system "represents" objective features of the external world. Such informal usages conform to Tyler Burge's starting point: "the premise that our perceptual experience represents or is about objects, properties, and relations that are *objective*", a premise that goes beyond an individualist-internalist approach (1988). But these usages are on a par with an astronomer warning that a comet is aiming directly toward the Earth, implying no animist, intentional physics.

The internalist study of language also speaks of "representations" of various kinds, including phonetic and semantic representations at the "interface" with other systems. But here too we need not ponder what is represented, seeking some objective construction from sounds or things. The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether the consequence of tachistoscopic presentations or of a real rotating cube or of stimulation of the retina in some other way; or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world, as might be suggested by a well-known philosophical tradition and inappropriate analogies from informal usage. Misperception raises no difficulties for this approach; it is a matter of how people assign interpretations to interactions they observe—to the reactions of a frog or person in an experiment, a photoreceptor that is "deceived", etc.—a fair topic for internalist inquiry into the psychology of the person who is deciding what to call a "misperception".

For psychology and ethnoscience, little seems at stake in these debates. Suppose Jones is a member of some ordinary community, and *J* is indistinguishable from him except that his total experience derives from some virtual reality design; or let *J* be Jones's Twin in a Twin-Earth scenario. They have had indistinguishable experiences and will behave the same way (insofar as behavior is predictable at all); they have the same internal states. Suppose that *J* replaces Jones in the community, unknown to

anyone except the observing scientist. Unaware of any change, everyone will act as before, treating *J* as Jones; *J* too will continue as before. The scientist seeking the best theory of all of this will construct a narrow individualist account of Jones, *J*, and others in the community. The account omits nothing, including the way members of the community attribute mental states (beliefs, meanings, perceptual contents, etc.), if they do.

Suppose that the community contains a philosopher *P* with the externalist intuitions of recent discussion. The theory will assign to *P* the corresponding internal state. It will now predict correctly that *P*, taking *J* to be Jones, will attribute to *J* the mental states he did to Jones; and that if aware of the *J*—Jones interchange when it occurs, *P* will attribute different mental states to *J*. Not sharing *P*'s intuitions, I don't know how *P* would attribute mental states as *J* lives on in the community, in a world of "objective" things (does *J* now come to share Jones's beliefs?). But whatever the answer, the theory will describe *P*'s internal states accordingly. If I am a member of the community too, the theory will assign to me a different internal state, in which no fixed answers are given about attribution of beliefs and meanings to *J* (and nothing interesting about contents, perceptual or other, because I take the technical innovations to mean what their designers say), various judgments being given as circumstances vary.

This account deals with Jones, *J*, other community members, and people with various intuitions about attribution of mental states; it is incomplete insofar as these intuitions are as yet unknown, but otherwise nothing seems missing from it, and it can readily be extended to the usage of other languages and cultures, as they differ. It can be converted easily enough into a non-individualist theory, more cumbersome and adding no new insight. That step would be inappropriate for naturalistic inquiry, and it is unclear what other purpose it might serve.

Talk about organs or organisms "solving problems", or being adapted to their functions, is to be understood similarly: as metaphoric shorthand. There is no question as to whether the wings of a butterfly are designed to "solve the problem" of flight; they evolved as thermoregulators, and still serve that purpose. If we were to learn that they reached their current state before they were ever used to fly, they would still now have the function of flight and would serve that purpose. The human visual system is maladapted to seeing in the dark, but is not a failure for that reason. The spine of large vertebrates is badly designed from an engineering standpoint, as most people know from their personal experience; but it is neither a success nor a failure. Human languages are in part unusable, but none the worse for that; people use the parts that are usable. It has very recently been discovered that while insects seem marvellously adapted to particular kinds

of flowering plants, in fact insects achieved virtually their present diversity and structure millions of years before flowering plants existed. When they appeared, “there was already waiting for them an encyclopedia of solutions waiting for the problems to be solved”, Richard Lewontin points out, intending to stress the meaninglessness of these intuitive categories for biology. It is, correspondingly, a misreading of informal talk to conclude that Marr’s theory of vision attributes “intentional states that represent objective, physical properties” because “There is no other way to treat the visual system as solving the problem that the theory sees it as solving” (Burge 1986). The theory itself has no place for the concepts that enter into the informal presentation, intended for general motivation. The statement “the idea that we classify our perceptual phenomenology without specifying the objective properties that occasion it is wildly out of touch with actual empirical theories of perception as well as with common sense” (Burge 1988) is correct in some circumstances with regard to common sense, but misleading with regard to empirical theories of perception, which are concerned with how things work and with perceptual reports and intuitive classifications only as evidence bearing on this matter.²⁵

Studying any organic system, a biologist naturally takes into account environmental interactions and physical law that are likely to have influenced mutations, reproductive success, and the course of development. For motivation and intuitive guidance, the biologist might speak of systems as having “evolved to solve certain problems forced on them by the environment”, with “Different species [set] different problems and solv[ing] them differently” (Burge 1988). But this is informal talk, and if it is discovered that the course of evolution was not what had been thought, as in the case of insects and flowers, the actual theory of sensory processing and other systems is not modified, with different attributions and individuation, and revised descriptions of intentional content, mistakes, functions, purposes, problems solved, and so on. Similarly, suppose it were discovered that our ancestors had been constructed in an extraterrestrial laboratory and sent to earth by space ship 30,000 years ago, so that natural selection played virtually no role in the formation of the kidney, visual system, arithmetical competence, or whatever. The technical sections of textbooks on the physiology of the kidney would not be modified, nor the actual theory of the functions computed by the retina or of other aspects of the human visual and other systems.

The critique of internalism (individualism) gains no more force from the observation that, in normal environments, internal processes are

²⁵ Lewontin (1994); Labandeira and Sepkoski (1993). The discussions in the literature about “what Marr meant” are somewhat strange; what matters is what a scientist does, not what he may have had in mind. For what seems to me an accurate account of Marr’s actual theory, see Frances Egan n.d..

reliably correlated with distal properties (object boundaries, and so on). In other environments, they correlate with different properties, which may be distal properties or direct retinal (or deeper internal) stimulation. We can say, if we like, that “where the constraints that normally enable an organism to compute a cognitive function are not satisfied, it will fail to represent its environment” (Egan n.d.); but that “failure” is our way of describing some human end that we impose for reasons unrelated to naturalistic inquiry, much as in the case of the failure of a comet to hit Jupiter, as we hoped it would. Nor is it relevant that consideration of “representation” in normal environments allows us to associate the system under analysis with the informally described cognitive function of vision. It’s no task of science to conform to the categories of intuition, or to decide whether it is still “vision” in abnormal environments or if parts of the brain normally used for other purposes take over some of the analysis of visual images, as they sometimes do. The study of perception naturally begins with informally presented “cognitive tasks”, but cares little whether something similar to them is discovered as it progresses.

Informal discussion of evolutionary processes makes use of such locutions as “solving problems”, but again that is not to be taken too seriously. Physical law provides narrow channels within which complex organisms may vary, and natural selection is doubtless a factor in determining the distribution of traits and properties within these constraints. A factor, not *the* factor, at least if we follow Darwin’s sensible strictures. Much concerned by the misinterpretation of his ideas, Darwin firmly denied that he attributed “the modification of species exclusively to natural selection”, emphasizing in the last edition of *Origin of Species* that

In the first edition of this work, and subsequently, I placed in a most conspicuous position—namely, at the close of the Introduction—the following words: “I am convinced that natural selection has been the main but not the exclusive means of modification”. This has been of no avail. Great is the power of steady misrepresentation.

Darwin took explicit note of a range of possibilities, including nonadaptive modifications and unselected functions determined from structure (see Gould 1982, p. 49-50).

We cannot sensibly estimate the weight that will be assigned to natural selection as a mechanism of evolution as more is learned about complex systems, the operation of physical law, the factors in spontaneous self-organization in living as in other physical systems, and so on.²⁶ The status

²⁶ See Waldrop (1990); Bradley (1994). The proposals reported in the latter review have been undermined, but the problem remains of accounting for prevailing asymmetries ranging from the “molecular handedness” of amino acids and DNA through location and orientation of organs.

of internalist approaches is unaffected by such considerations, whether we are thinking of ants and the kidney, or language and mind.

Virtually every aspect of the study of language and mind seems to me to involve unjustified non-naturalist assumptions.²⁷ If this discussion is on the right track, one would want to ask why such ideas appear so compelling. The answer could be that our common sense picture of the world is profoundly dualistic, ineradicably, just as we can't help seeing the setting of the sun, or sharing Newton's belief in the "mechanical philosophy" that he undermined, or watching the wave that "flees the place of its creation", as Leonardo put it, independently of what we may know in some other corner of our minds. If so, and if metaphysical dualism has been undermined, what is left is a kind of methodological dualism, an illegitimate residue of common sense that should not be allowed to hamper efforts to gain understanding into what kind of creatures we are.

Department of Linguistics and Philosophy
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

NOAM CHOMSKY

REFERENCES

- Atlas, J.D. 1989: *Philosophy Without Ambiguity*. Oxford: Clarendon Press.
- Baker, Lynne Rudder 1988: "Cognitive Suicide", in R.H. Grimm and D.D. Merrill, eds., *Contents of Thought*. Tucson: University of Arizona Press, pp. 1–18.
- Baldwin, Thomas 1993: "Two Types of Naturalism", *Proceedings of the British Academy*, 80, pp. 171–99.
- Barinaga Marcia 1994: "Neurons Tap Out a Code That May Help Locate Sounds". *Science*, 264, p. 775.
- Bilgrami, Akeel 1992: *Belief and Meaning*. Oxford, Basil Blackwell.
- Bradley, D. 1994: "Organic Chemistry: A New Twist in the Tale of Nature's Asymmetry". *Science*, 264, p. 908.
- Brock, William H. 1992: *The Fontana/Norton History of Chemistry*. New York, London: Norton.
- Bromberger, Silvain 1992: "Types and Tokens in Linguistics", in Bromberger, *On What We Know We Don't Know*. Chicago: University of Chicago Press.
- Burge, Tyler 1986: "Individualism and Psychology". *Philosophical Review*, 95, pp. 3–45.

²⁷ For more extensive discussion, see Chomsky (forthcoming).

- 1988: “Cartesian Error and the Objectivity of Perception”, in R.H. Grimm and D.D. Merrill, eds., *Contents of Thought*. Tucson: University of Arizona Press, pp. 62–98.
- 1992: “Philosophy of Language and Mind”. *Philosophical Review*, 101, pp. 3–52.
- Chastain, Charles 1988: “Comments on Baker”, in R.H. Grimm and D.D. Merrill, eds., *Contents of Thought*. Tucson: University of Arizona Press, pp. 18–26.
- Chomsky, Noam 1957: *Syntactic Structures*. The Hague: Mouton.
- 1966: *Cartesian Linguistics*. New York: Harper & Row.
- 1975a: *Reflections on Language*. New York: Pantheon.
- 1975b: *Logical Structure of Linguistic Theory*. New York: Plenum; University of Chicago Press. Original ms. 1955–6.
- 1980: *Rules and Representations*. New York: Columbia University Press.
- 1986: *Knowledge of Language*. New York: Praeger.
- 1992: “Language and Interpretation: Philosophical Reflections and Empirical Inquiry”, in John Earman, ed., *Inference, Explanation and other Philosophical Frustrations*. Berkeley, University of California Press.
- 1993a: “Explaining Language Use”, in James E. Tomberlin, (ed.), *Philosophical Topics*, 20, pp. 205–31.
- 1993b: “A Minimalist Program for Linguistic Theory”, in K. Hale and S.J. Keyser, eds., *The View From Building 20*. Cambridge: MIT Press.
- 1994: “Bare Phrase Structure”. *MIT Occasional Papers in Linguistics*, 5.
- forthcoming: “Naturalism and Dualism in the Study of Language and Mind”, *Irish Journal of Philosophy*.
- Chomsky et al. 1993: *Language and Thought*. London, Warwick, Rhode Island: Moyer Bell.
- Churchland, Patricia 1994: “Can Neurobiology Teach Us Anything about Consciousness?”, *Proceedings and Addresses of the American Philosophical Association*, 67, 4, pp. 23–40.
- Churchland, Paul 1994: Review of Searle 1992. *London Review of Books*, 12 May, p. 13–14.
- Clark, Andy and Karmiloff-Smith, Annette 1993: “The Cognizer’s Innards”. *Mind and Language*, 8.4, pp. 487–519.
- Davidson, Donald 1986: “A Nice Derangement of Eptiaphs”, in E. Lepore, (ed.), *Truth and Interpretation*. Oxford: Blackwell.
- 1990: “The Second Person”, ms. University of California at Berkeley.

- Dennett, Daniel 1991: Review of McGinn 1991. *Times Literary Supplement*, 4597, p. 10.
- Devitt, Michael and Sterelny, Kim 1989: "Linguistics: What's Wrong with 'The Right View'". *Philosophical Perspectives*, 3, pp. 497–531.
- Dijksterhuis, E. J. 1986: *Mechanization of the World Picture*. Princeton: Princeton University Press.
- Dreben, Burton 1993: "Putnam, Quine—and the Facts", in James E. Tomberlin, (ed.), *Philosophical Topics*, 20, pp. 293–315.
- Dummett, Michael 1986: "Comments on Davidson and Hacking", in E. LePore 1986, pp. 459–476.
- 1991: *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.
- 1993: *The Seas of Language*. Oxford: Oxford University Press.
- Edelman, Gerald 1992: *Bright Sun, Brilliant Fire*. New York: Basic Books.
- Egan, Frances n.d.: "Computation and Content", ms. Rutgers University.
- Friedman, Michael 1993: "Remarks on the History of Science and the History of Philosophy", in P. Horwich, ed., *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge: MIT Press.
- Gay, Peter 1977: *The Enlightenment: An Interpretation*, Vol. II. New York: Norton.
- Gleitman, Lila 1990: "The Structural Sources of Verb Meanings". *Language Acquisition: A Journal of Developmental Linguistics*, 1, pp. 3–55.
- Gould, Stephen 1982: *The Panda's Thumb*. New York: Norton.
- Haugeland, John 1979: "Understanding Natural Language". *Journal of Philosophy*, 76, pp. 519–32.
- Hagoort, Peter et al. 1992: "The Syntactic Positive Shift (SPS) as an ERP-measure of Syntactic Processing", ms. Max Planck Institute for Psycholinguistics, Nijmegen, to appear in *Language and Cognitive Processes*.
- Hagoort, Peter and Brown, Colin 1993: "Brain Responses to Lexical Ambiguity, Resolution, and Parsing", ms. Max Planck Institute, to appear in C. Clifton et al., eds., *Perspectives on Sentence Processing*. Hillsdale, New Jersey: Erlbaum.
- Jacob, François 1973: *The Logic of Life*. New York: Pantheon.
- Jacob, Margaret 1988: *The Cultural Meaning of the Scientific Revolution*. Philadelphia: Temple University Press.
- Labandeira, C.C. and Sepkoski, J.J. 1993: "Insect Diversity in the Fossil Record". *Science*, 261, pp. 310–15.
- LePore, E., ed. 1986: *Truth and Interpretation*. Oxford: Basil Blackwell.
- Lewontin, Richard 1994: *La Repubblica*. ms. Harvard

- Llinás, Rodolfo 1987: "Mindness as a Functional State of the Brain", in Colin Blakemore and Susan Greenfield, eds., *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*. Oxford: Basil Blackwell.
- Marr, David 1982: *Vision*. San Francisco: W.H. Freeman.
- Marshall, John 1989: "On Making Representations", in C. Brown, P. Hagoort and T. Meijering, eds., *Vensters op de geest*. Utrecht: Stichting Grafiet.
- McGinn, Colin 1991: *The Problem of Consciousness*. Oxford: Basil Blackwell.
- 1993: *Problems in Philosophy*. Oxford: Basil Blackwell.
- Mehler, Jacques and Dupoux, Emmanuel 1994: *What Infants Know*. Oxford: Basil Blackwell.
- Middlebrooks, J.C., Clock A.E., Green, D.M. and Xu, Li 1994: "A Panoramic Code for Sound Location by Cortical Neurons". *Science*, 264, pp. 842-44.
- Millikan, Ruth 1984: *Language, Thought, and Other Biological Categories*. Cambridge MA: MIT Press.
- Moravcsik, Julius 1975: "Aitia as Generative Factor in Aristotle's Philosophy". *Dialogue*, 14, pp. 622-36.
- 1990: *Thought and Language*. London: Routledge.
- Nagel, Thomas 1993: "The Mind Wins!", review of Searle 1992. *New York Review of Books*, pp. 37-41.
- Neville, Helen et al. 1991: "Syntactically Based Sentence Processing Classes: Evidence from Event Related Brain Potentials". *Journal of Cognitive Neuroscience*, 3, pp. 151-65.
- Passmore, John 1965: *Priestley's Writings on Philosophy of Science and Politics*. New York, London: Collier-MacMillan.
- Pustejovsky, James, ed. 1993a: *Semantics and the Lexicon*. Dordrecht: Kluwer.
- 1993b: "Type Coercion and Lexical Selection", in Pustejovsky 1993a. Dordrecht: Kluwer.
- Putnam, Hilary 1975: "The Meaning of Meaning", in Keith Gunderson, ed., *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science, Vol. VII. Minneapolis: University of Minnesota Press.
- 1989: *Representation and Reality*. Cambridge: MIT Press.
- 1993: "Replies", in James E. Tomberlin, ed., *Philosophical Topics*, 20, pp. 347-408.
- Quine, W.V. 1992: "Structure and Nature". *Journal of Philosophy*, 89, pp. 5-9.

- Romaine, Suzanne 1994: *Language in Society*. Oxford: Oxford University Press.
- Schweber, Silvan S. 1993: "Physics, Community and the Crisis in Physical Theory". *Physics Today*, 46, 11, pp. 34–40.
- Searle, John 1992: *The Rediscovery of Mind*. Cambridge: MIT Press.
- Smith, Barry 1992: "Understanding Language". *Proceedings of the Aristotelian Society*, pp. 109–41.
- Smith, Neil, Tsimpli, Ianthi-Maria, and Ouhalla, Jamal 1993: "Learning the Impossible: The Acquisition of Possible and Impossible Languages by a Polyglot Savant". *Lingua*, 91, pp. 279–347.
- Soames, Scott 1989: "Semantics and Semantic Competence". *Philosophical Perspectives*, 3, pp. 575–96.
- Stich, Stephen P. 1983: *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.
- Stryker, Michael 1994: "Precise Development from Imprecise Rules". *Science*, 263, pp. 1244–5.
- Turing, Alan 1950: "Computing Machinery and Intelligence". *Mind*, 59, 236, pp. 433–460.
- Ullman, S. 1979: *The Interpretation of Visual Motion*. Cambridge MA: MIT Press.
- Waldrop, M.M. 1990: "Spontaneous Order, Evolution, and Life". *Science*, 247, pp. 1543–4.
- Wright, Crispin 1989: "Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics", in A. George, ed., *Reflections on Chomsky*. Oxford: Basil Blackwell, pp. 233–64.
- Yamada, Jeni 1990: *Laura*. Cambridge: MIT Press.
- Yolton, John 1983: *Thinking Matter*. Minneapolis: University of Minnesota Press.