

## 4 Naturalism and dualism in the study of language and mind

---

The terms of the title can be understood in various ways, along with the frameworks in which they are embedded. I would like to outline interpretations that I think are useful and proper, and to suggest a more general thesis, which would require much more comprehensive argument: that there is no coherent alternative to proceeding in this way for the range of issues addressed, and that other endeavors in roughly the same realm are clarified and facilitated if understood as extensions of the approach outlined.

### **Deflating the terms**

Putting “language” aside for the moment, let’s begin by taking the other terms of the title in ways that are innocent of far-reaching implications, specifically, divorced from any metaphysical connotations. Take the term “mind” or, as a preliminary, “mental.” Consider how we use such terms as “chemical,” “optical,” or “electrical.” Certain phenomena, events, processes, and states are called “chemical” (etc.), but no metaphysical divide is suggested by that usage. These are just various aspects of the world that we select as a focus of attention for the purposes of inquiry and exposition. I will understand the term “mental” in much the same way, with something like its traditional coverage, but without metaphysical import and with no suggestion that it would make any sense to try to identify the true criterion or mark of the mental. By “mind,” I mean the mental aspects of the world, with no concern for defining the notion more closely and no expectation that we will find some interesting kind of unity or boundaries, any more than elsewhere; no one cares to sharpen the boundaries of “the chemical.”

Furthermore, I keep here to the human mind (visual system, reasoning, language, etc.). There is no quest for a unified science of locomotion, ranging from amoeba to eagle to science-fiction spaceship; or of communication, ranging from cell to poetic discourse to imagined extraterrestrials. Rather, biologists study how dolphins swim and ants communicate,

beginning with an “internalist” and “individualist” account (in contemporary jargon). In so doing, they have little interest in how the terms “dolphin,” “communicate,” etc. are used in the informal discourse in which the questions are initially posed. Rather, they develop concepts appropriate to their purpose of explanation and understanding. Ordinary discourse and common-sense thought are in no way denigrated by the procedure; rather they are liberated from inappropriate and destructive demands. The same is true of other scientific inquiry with broader concerns (for example the study of ant communities).<sup>1</sup>

We may carry over these observations – truisms, I think – to the study of human language and the human mind. Since the brain, or elements of it, are critically involved in linguistic and other mental phenomena, we may use the term “mind” – loosely but adequately – in speaking of the brain, viewed from a particular perspective developed in the course of inquiry into certain aspects of human nature and its manifestations. There are empirical assumptions here – that the brain, not the foot, is the relevant bodily organ, that humans are alike enough in language capacity so that human language can be regarded as a natural object, and so on. But these need not detain us.

Let us also understand the term “naturalism” without metaphysical connotations: a “naturalistic approach” to the mind investigates mental aspects of the world as we do any others, seeking to construct intelligible explanatory theories, with the hope of eventual integration with the “core” natural sciences. Such “methodological naturalism” can be counterposed to what might be called “methodological dualism,” the view that we must abandon scientific rationality when we study humans “above the neck” (metaphorically speaking), becoming mystics in this unique domain, imposing arbitrary stipulations and *a priori* demands of a sort that would never be contemplated in the sciences, or in other ways departing from normal canons of inquiry.

There are interesting questions as to how naturalistic inquiry should proceed, but they can be put aside here, unless some reason is offered to show that they have a unique relevance to this particular inquiry. That has not been done, to my knowledge. Specifically, skeptical arguments can be dismissed in this context. We may simply adopt the standard outlook of modern science, in essence, the anti-foundationalism of the seventeenth century reaction to the Cartesian skeptical crisis, as Richard Popkin describes it: “the recognition that absolutely certain grounds could not be given for our knowledge, and yet that we possess standards for evaluating the reliability and applicability of what we have found out about the world,” thus “accepting and increasing the knowledge itself”

while recognizing that “the secrets of nature, of things-in-themselves, are forever hidden from us” (Popkin 1979: 139ff.). It may well be of interest to proceed beyond but, if so, the place to look for answers is where they are likely to be found: in the hard sciences, where richness and depth of understanding provides some hope of gaining insight into the questions. To raise them with regard to inquiries barely attempting to gain a foothold is pointless, scarcely more than a form of harassment of emerging disciplines.

Naturalism, so understood, should be uncontroversial, though its reach remains to be determined; and the dualistic alternative should be highly controversial. I think that the opposite has been true, a curious feature of recent intellectual history. Explanatory theories of mind have been proposed, notably in the study of language. They have been seriously challenged, not for violating the canons of methodological naturalism (which they seem to observe, reasonably well), but on other grounds: “philosophical grounds,” which are alleged to show that they are dubious, perhaps outrageous, irrespective of success by the normal criteria of science; or perhaps that they are successful, but do not deal with “the mind” and “the mental.” I will suggest that such critiques are commonly a form of methodological dualism, and that advocacy (or tacit acceptance) of that stance has been a leading theme of much of the most interesting work in recent philosophy of mind and language.

Plainly, a naturalistic approach does not exclude other ways of trying to comprehend the world. Someone committed to it can consistently believe (I do) that we learn much more of human interest about how people think and feel and act by reading novels or studying history or the activities of ordinary life than from all of naturalistic psychology, and perhaps always will; similarly, the arts may offer appreciation of the heavens to which astrophysics does not aspire. We are speaking here of theoretical understanding, a particular mode of comprehension. In this domain, any departure from this approach carries a burden of justification. Perhaps one can be given, but I know of none.

### **Language in naturalistic inquiry**

To help frame the discussion, let’s consider for a moment where methodological naturalism leads us in the study of mind, language in particular. I think to something like the following, on current understanding.

The brain has a component – call it “the language faculty” – that is dedicated to language and its use. For each individual, the language faculty has an initial state, determined by biological endowment. Serious

pathology apart, such states are so similar across the species that we can reasonably abstract to *the* initial state of the language faculty, a common human possession. The environment triggers and to a limited extent shapes an internally-directed process of growth, which stabilizes (pretty much) at about puberty. A serious study will attempt to determine what “pure” states of the language faculty would be under ideal conditions, abstracting from a host of distortions and interferences in the complex circumstances of ordinary life, thus hoping to identify the real nature of the language faculty and its manifestations; at least, so the canons of methodological naturalism dictate. This point of view, adopted without comment in naturalistic inquiry generally, is often considered contentious or worse in the domain of language and mind, an illustration of the dualism that I suggested is prevalent and pernicious.

A state attained by the language faculty characterizes an infinite class of linguistic expressions, each a certain array of phonetic, structural, and semantic properties. My state specifies the properties of the last sentence; yours is similar enough so that your mind can (sometimes) find an appropriate analogue to what I say, in which case you have means for determining my intentions (the perceived expression being only part of your evidence, and communication being a “more or less” affair). The state attained is a computational (generative) system. We may call that state a language or, to avoid pointless terminological controversy, an *I-language*, “I” chosen to suggest that the conception is internal, individual, and intensional (in the technical sense; that is, the characterization of a function in intension). For Jones to have the (I-)language, L, is for his language faculty to be in state L. Particular signals are manifestations of linguistic expressions (spoken, written, signed, whatever); speech acts are manifestations of linguistic expressions in a broader sense. The expressions can be understood as “instructions” to other systems of the mind/brain that “follow them” in the use of language.

On the (very weak) empirical assumptions of these remarks, the notion I-language is straightforward; that the brain is a complex system with states and properties is not controversial. It remains to spell out this conception of “state of the brain” and to discover its properties. Other notions of “language” require some further justification – which, I believe, is not easy to give.

The class of expressions generated by the (I-)language L should not be confused with a category of well-formed sentences, a notion that has no known place in the theory of language, though informal exposition has sometimes obscured the point, leading to much confusion and wasted effort. Thus, so-called “deviant” expressions may be characterized

by Jones's language *L* with quite definite properties; it could turn out that it assigns a specific interpretation to every possible signal, the latter notion determined by properties of the initial state.

It may be that the computational system itself is (virtually) invariant, fixed by innate biological endowment, variation among languages and language types being limited to certain options in the lexicon; quite restricted options. Slight changes in an intricate system may yield what appear to be dramatic phenomenal differences; thus, languages may appear to differ radically from one another, though they differ only in rather marginal ways, it appears. Something like that is what any rational scientist observing humans would expect; otherwise, there would be little hope of accounting for the specificity, richness, and intricacy of the state attained on the basis of very limited information from the environment. Comparable assumptions are taken for granted without discussion in the study of growth and development generally. A naturalistic approach makes no distinction in the unique case of mental processes.

As far as is known, even the most rudimentary properties of the initial and attained states are not found among other organisms or, indeed, in the biological world, apart from its points of contact with inorganic matter. Nor are there more than very weak relations to anything discovered in the brain sciences. So we face the problems of unification that are common in the history of science, and do not know how – or if – they will be resolved.

I'll put aside here any further account of the results of naturalistic inquiry, returning to the questions of naturalism and dualism more generally.

### Varieties of naturalism

Methodological naturalism is not to be confused with other varieties. To clarify what I do and don't mean, consider a useful recent exposition of the concept of naturalism by Baldwin (1993: 171). He opens by noting that "A prominent theme of current philosophy is that of the 'naturalisation' of philosophy. Daniel Dennett has written that 'One of the happiest trends in philosophy in the last twenty years has been its Naturalisation'" (p. 171). That the trend is prominent is doubtless true; that it is happy seems to me open to question. In any event, it is distinct from the form of naturalism I am advocating here.

Baldwin finds "two different types of naturalism at work in current philosophy," what he calls *metaphysical* and *epistemic*. The former is what "Dennett has in mind when he celebrates the 'naturalisation' of

philosophy”: the thought that, as Dennett puts it, “philosophical accounts of our minds, our knowledge, our language must in the end be continuous with, and harmonious with, the natural sciences” (p. 172) – unlike, say, Fregean Platonism, which is not continuous with hypotheses “advanced by the natural sciences,” so it is alleged.

Contemporary epistemic naturalism derives from Willard Quine’s “epistemology naturalized,” which stipulated that the study of knowledge and belief must be incorporated within a narrow branch of behaviorist psychology of no known scientific interest, a strange move in itself, which has evoked surprisingly little challenge. A broader version, Baldwin observes, considers “natural relations” between external situations and mental states without arbitrary strictures. The broader version can be viewed as an outgrowth of the rational psychology of the seventeenth century, which held, as Lord Herbert put it, that there are “principles or notions implanted in the mind” that “we bring to objects from ourselves . . . [as] . . . a direct gift of Nature, a precept of natural instinct” – “common notions” and “intellectual truths” that are “imprinted on the soul by the dictates of Nature itself,” which, though “stimulated by objects,” are not “conveyed” by them (Herbert 1624/1937: 133). Baldwin cites Thomas Reid as the source of a kind of “naturalised epistemology,” expressing a similar point of view but “freed from Hume’s [or any earlier] commitment to the theory of ideas” (Baldwin 1993: 181); that is, freed from earlier attempts to spell out what Reid calls the “original and natural judgments” that “nature hath given to the human understanding” as “part of our constitution” and that make up “*the common-sense of mankind*” (Reid 1785: 600–1). Since nothing replaces the outline of a theory that is abandoned, it is hard to see how this “naturalization” progresses beyond earlier versions. On the contrary, the work of the Cartesians and Cambridge Platonists is considerably more advanced in many respects, in my opinion. Later, Charles Sanders Peirce (1957: 253) proposed that human thought is guided by a principle of “abduction” that “puts a limit upon admissible hypotheses” and that is innate in us, providing the human mind with “a natural adaptation to imagining correct theories of some kind” (p. 238) a result of natural selection, he suggested (with little plausibility). There are many further ramifications, including recent “evolutionary epistemology.” (For some discussion, see Chomsky 1966: Chapter 4; 1968/72; 1975: Chapter 1.)

The enterprise of epistemic naturalism is uncontentious, apart from the term, which is misleading in a peculiarly modern way. The epistemic naturalism of the seventeenth and eighteenth century was science, an attempt to construct an empirical theory of mind; Hume, for one, compared his enterprise with Isaac Newton’s. Epistemic naturalism, in

contrast, is presented as a “philosophical position,” something apparently different. We plainly cannot read back into earlier periods a distinction between science and philosophy that developed later. We would not use the term “visual naturalism” to refer to the empirical study of the growth and functioning of the visual system (also a topic of earlier rational psychology), implying that there was some coherent alternative for the same realm of problems. The term “epistemic naturalism” seems to me misleading in much the same way, not to speak of the special versions deriving from Quine’s “epistemology naturalized.”

For a methodological naturalist, traditional epistemic naturalism is normal science (see Chapter 3 of this volume, pp. 52–3), however we evaluate particular implementations. Inquiry into the initial state of the language faculty, for example, is an attempt to discover the “principles or notions implanted in the mind” that are a “direct gift” of nature, that is, our biological endowment. As elsewhere, the inquiry is initiated by common-sense formulations. Take the informal locution “Jones knows (speaks, understands, has) English.” The observation focuses attention on a state of the world, including a state of Jones’s brain, a cognitive state, that underlies Jones’s knowledge of many particular things: his knowing how to interpret linguistic signals, or that certain expressions mean what they do, and so on. We would like to know how Jones’s brain reached this cognitive state. Inquiry into the matter leads to empirical hypotheses about biological endowment, interactions with the environment, the nature of the states attained, and their interactions with other systems of the mind (articulatory, perceptual, conceptual, intentional, etc.). Resulting theories of the growth of language are sometimes called theories of a “Language Acquisition Device” (LAD), which effects a transition from the initial state of the language faculty to later states, mapping experience to state attained; the theory of the initial state is sometimes called “Universal Grammar” (UG), adapting a traditional notion to a somewhat different context. (I ignore below the distinctions between the theory of LAD and UG.) In my terms, this is study of the mind; others disagree, for reasons to which I will return.

Metaphysical naturalism seems far more problematic than traditional epistemic naturalism. One question, which Baldwin raises, is “what the ‘natural’ sciences are.” A possible answer is: whatever is achieved in pursuing naturalistic inquiry. But that doesn’t seem to be what is intended; let us put the question to the side for a moment. A related problem is to explain what are “philosophical accounts of our minds, our knowledge, our language,” and how they differ from “scientific accounts,” particularly if they are “continuous with the natural sciences” (Baldwin 1993: 172). Does the doctrine mean that a theory of

mind should be “continuous” and “harmonious” with today’s physics? That is surely unacceptable; tomorrow’s physics may well not meet that condition. With some Peircean ideal of what science will be “in the limit”? Not very helpful, even if meaningful. Perhaps tomorrow’s physics will incorporate some version of today’s accounts (whether termed “philosophical” or not), even if the latter are not continuous with today’s physics.

If so, it will be nothing new in the history of the sciences. One persistent goal is to unify various theories about the world, but the process has taken many a different course. Large-scale reduction is not the usual pattern; one should not be misled by such dramatic examples as the reduction of much of biology to biochemistry in the middle of the twentieth century. Repeatedly, the more “fundamental” science has had to be revised, sometimes radically, for unification to proceed. Suppose that a nineteenth century philosopher had insisted that “chemical accounts of molecules, interactions, properties of elements, states of matter, etc. must in the end be continuous with, and harmonious with, the natural sciences,” meaning physics as then understood. They were not, because the physics of the day was inadequate. By the 1930s, physics had radically changed, and the accounts (themselves modified) were “continuous” and “harmonious” with the new quantum physics. Suppose that a seventeenth-century scientist were to have imposed the same demand on celestial mechanics, referring to the prevailing “mechanical philosophy” and rejecting Newton’s mystical theory (as Leibniz and Huygens did), because it was incompatible with “the laws of mechanics.” (See Dijksterhuis 1986: 479f.) Though understandable, the reaction would have been (and was) surely wrong: fundamental physics had to be radically changed for unification to proceed.

We have no idea where that process will lead, or even how far human intelligence can reach in attaining such understanding of the natural world; we are, after all, biological organisms, not angels. The latter observation, again uncontentious, suggests another way to answer the question of “what the ‘natural’ sciences are.” Among the aspects of the mind are those that enter into naturalistic inquiry; call them “the science-forming faculty” (SFF). Equipped with SFF, people confront “problem situations,” consisting of certain cognitive states (of belief, understanding, or misunderstanding), questions that are posed, and so on (essentially, what Sylvain Bromberger calls a “p-predicament”; see his essays collected in Bromberger 1992b). Often SFF yields only a blank stare. Sometimes it provides ideas about how the questions might be answered or reformulated, or the cognitive state modified, ideas that can then be evaluated in ways that SFF offers (empirical test, consistency



with other parts of science, criteria of intelligibility and elegance, etc.). Like other biological systems, SFF has its potential scope and limits; we may distinguish between *problems* that in principle fall within its range, and *mysteries* that do not. The distinction is relative to humans; rats and Martians have different problems and mysteries and, in the case of rats, we even know a fair amount about them. The distinction also need not be sharp, though we certainly expect it to exist, for any organism and any cognitive faculty. The successful natural sciences, then, fall within the intersection of the scope of SFF and the nature of the world; they treat the (scattered and limited) aspects of the world that we can grasp and comprehend by naturalistic inquiry, in principle. The intersection is a chance product of human nature. Contrary to speculations since Peirce, there is nothing in the theory of evolution, or any other intelligible source, that suggests that it should include answers to serious questions we raise, or even that we should be able to formulate questions properly in areas of puzzlement.

Specifically, it is unknown whether aspects of the theory of mind – say, questions about consciousness – are problems or mysteries for humans, though in principle we could discover the answer, even discover that they are mysteries; there is no contradiction in the belief that SFF might permit us to learn something about its limits. (See Chomsky, 1968 ch. 3; 1975, ch. 4. On the possible limits, and the relevance to philosophical inquiry, see particularly McGinn 1991; 1993.)

The question “what the ‘natural’ sciences are,” then, might be answered narrowly, by asking what they have achieved; or more generally, by inquiry into a particular faculty of (the human) mind, with its specific properties. Something else, however, seems to be wanted; what it is remains unclear.

It is instructive to look more closely at the origins of modern science. In brief, progress into the seventeenth century laid the basis for the “mechanical philosophy,” eliminating fantasies about forms of objects floating through the air and implanting themselves in brains, and mystical forces and powers, “occult qualities” of sympathy, antipathy, and so on, which allowed such absurdities as action at a distance through a vacuum. The Cartesians observed that certain phenomena of nature (notably, the normal use of language) did not seem to fall within the mechanical philosophy, postulating a new principle to account for them. Given their metaphysics, they postulated a second substance (*res cogitans*, mind), for other reasons as well. Implementation aside, the move was not unreasonable, in fact, not unlike Newton’s reasoning when he discovered the inadequacies of the mechanical philosophy. Postulation of something that lies beyond the mechanical philosophy gives rise to two

tasks: to develop the theory and to solve the unification problem; in the Cartesian case, the “mind–body problem.” All of this is normal science; wrong, but that is also the norm.

Just as the mechanical philosophy appeared to be triumphant, it was demolished by Newton, who reintroduced a kind of “occult” cause and quality, much to the dismay of leading scientists of the day, and of Newton himself. The Cartesian theory of mind (such as it was) was unaffected by his discoveries, but the theory of body was demonstrated to be untenable. To put it differently, Newton eliminated the problem of “the ghost in the machine” by exorcising the machine; the ghost was unaffected. He also left us with the conclusion that common-sense intuition – the “folk physics” that was the basis for the mechanical philosophy – cannot be expected to survive the transition to rational inquiry into the nature of things. The mind–body problem disappeared, and can be resurrected, if at all, only by producing a new notion of body (material, physical, etc.) to replace the one that was abandoned; hardly a reasonable enterprise, it would seem. Lacking that, the phrase “material” (“physical,” etc.) would simply offer a loose way of referring to what we more or less understand and hope to unify in some way.

The natural conclusion, drawn shortly after by La Mettrie and later Joseph Priestley, is that human thought and action are properties of organized matter, like “powers of attraction and repulsion,” electrical charge, and so on. (La Mettrie 1747; see also Cohen 1941; Yolton 1983; Wellman 1992.) Adopting that view, we seek to determine the properties of these things in the world and to account for mental phenomena in terms of them, to show how they arise in the individual and species, and to relate these conclusions to whatever else is known about organized matter (the new version of the unification problem). On the last problem, there is no progress to speak of. Nor has there been real progress in accounting for the properties of normal use of language, and other phenomena, that led the Cartesians to postulate a second substance (though the limits of mechanism are no longer an issue). These may well turn out to be mysteries-for-humans. There has been progress in understanding the mechanisms of mind from the more abstract point of view of UG, LAD, the states attained, and their interactions with other cognitive systems; and in the study of some of these (for example, conceptual development). On naturalistic assumptions, these are parts of the natural sciences – good or bad, right or wrong.

The natural sciences attempt to understand the world in its chemical, electrical, mental, etc. aspects. Does the world include mysterious Newtonian forces affecting bodies separated by empty space, or electrical and magnetic fields that, though mathematical objects, are “*real* physical

‘stuff’” because of the way they “push each other along through empty space” (Penrose 1989: 185–6). Or curved space that “seemed to take all definite structure away from anything we can call solidity,” or perhaps, “at a very deep bottom” nothing but bits of information (Wheeler 1994: 294f.). Does it include Herbert’s common notions and principles as part of “natural instinct,” Humean ideas, thoughts and concepts, computational principles and states, and so on? Naturalistic inquiry seeks answers to these questions, as self-critically as it can, escaping arbitrary assumptions when these can be detected, though aware that biological constraints on human thought cannot be overcome, while cultural ones may not be easy to unravel.

Let us return to the allegation that a theory of mind, TM, that introduces such notions as “grasping Fregean senses” is not harmonious with or continuous with hypotheses “advanced by the natural sciences.” If one means the natural sciences of today, excluding TM, then the observation is correct though uninteresting. The right questions have to do with the status of TM on naturalistic grounds, and the unification problem (if TM has some plausibility). If the allegation means that the unification problem lies beyond human capacity, that could be right, but would not bear on the scientific status of TM. We need not consider speculations about the “true” science, perhaps beyond human intellectual reach. What else does metaphysical naturalism demand? That is not clear.

Shall we understand metaphysical naturalism to be the demand for unity of nature? If so, it could be taken as a guiding idea, but not as a dogma. “Ninety percent of the matter of the universe,” physicists tell us, “is what is now called dark matter – dark because we don’t see it; dark because we don’t know what it is,” indeed “we do not have the slightest idea of what 90 percent of the world is made of.” (Weisskopf 1989). Suppose dark matter turns out to be crucially different from the 10 per cent of the world about which there are some ideas. The possibility cannot be discounted in principle; stranger things have been accepted in modern science. Nor can it be excluded in the case of theories of mind. Though there is no reason to entertain the hypothesis, some version of Cartesianism (with a far richer concept of body) could in principle turn out to be true, consistent with a naturalist stance.

### **Materialism and its critics**

Metaphysical naturalism will be a coherent position if its advocates tell us what counts as “physical” or “material.” Until that is done, we cannot comprehend the doctrine, let alone such derivative notions as “eliminative materialism” and the like. In practice, versions of the latter

seem to be little more than pronouncements as to where the answers lie and, as such, are of no special interest.

Critics of these doctrines seem to me to be faced with the same problem: what are they criticizing? One of the most prominent is Thomas Nagel, who gives a lucid account of prevailing views and his critique of them, directed specifically to the questions that concern me here (Nagel 1993). I think the issues are wrongly put, though in an interesting way, and the conclusions suspect for this and other reasons, including those on LAD and the theory of mind, with which he concludes.

Nagel states that “the mind–body problem was posed in its modern form only in the seventeenth century, with the emergence of the scientific conception of the physical world on which we are now all brought up” (1993: 97) (the Newtonian conception). But that has the story reversed. The mind–body problem made sense in terms of the mechanical philosophy that Newton undermined, and has not been coherently posed since. If so, discussion cannot proceed in Nagel’s terms without some new account of the nature of body (material, physical, etc.) and mind.

This perspective on the issues and their origins leads to a misleading account of current contributions as well. Thus Nagel outlines John Searle’s “radical thesis” that “consciousness is a physical property of the brain” that is “irreducible to any *other* physical property, a position which, if properly clarified (which Nagel considers unlikely), “would be a major addition to the possible answers to the mind–body problem” (1993: 103). This thesis is the “metaphysical heart” of Searle’s proposal: in his own words, “consciousness is a higher-level or emergent property of the brain”; it is “as much of the natural biological order as . . . photosynthesis, digestion, or mitosis.”

Valid or not, the thesis is not radical; rather, it is – and was – the natural reaction to Newton’s demolition of the mechanical philosophy, hence of the mind–body problem, at least in its Cartesian form. As noted, the view that thought and action (including consciousness) are properties of organized matter, no more reducible to others than electromagnetic properties are reducible to mechanics, was put forth by eighteenth century scientists – not, however, as a possible answer to the mind–body problem, which had (and has) no coherent formulation. As for the metaphysical import of the thesis, it is on a par with the import of the relation between classical mechanics and electromagnetic theory.

Nagel assumes a prior understanding of mind and body, mental and physical, and gives some indication of what he means. Expressing a standard view, he takes “the essence of mind” to be consciousness: “all mental phenomena are either actually or potentially conscious” (1993: 97). Whether intended as a terminological or substantive proposal,

that formulation requires an explanation of the notion “potentially conscious”; Nagel adopts Searle’s (1992) proposal on the matter, but it seems to face serious difficulties.

Suppose we take consciousness to be the mark of the mental. What about body? That Nagel identifies with what is “capable of description by physical science” (excluding consciousness, whether by fiat or discovery is not clear). Thus he understands materialism (which he says is accepted by most contemporary philosophers) to be the belief “that everything there is and everything that happens in the world must be capable of description by physical science” – a view that he takes to be coherent, but false. Adopting it, one attempts “some sort of reduction of the mental to the physical – where the physical, by definition, is that which can be described in nonmental terms” (that is, terms that do not involve “potential consciousness”). “What is needed to complete the materialist world picture is some scheme of the form, ‘mental phenomena – thoughts, feelings, sensations, desires, perceptions, etc. – are nothing but . . . ,’ where the blank is to be filled in by a description that is either explicitly physical or uses only terms that can apply to what is entirely physical,” or perhaps gives “assertability conditions” on “externally observable grounds.” “The various attempts to carry out this apparently impossible task,” Nagel continues, “and the arguments to show that they have failed, make up the history of the philosophy of mind during the past fifty years.” Left unresolved, and presumably unresolvable, is the mind–body problem, which is the problem of “finding a place in the world for our minds themselves, with their perceptual experiences, thoughts, desires, scientific theory-construction, and much else that is not described by physics.”

The belief that the questions are coherent and significant is widely shared. Thus, 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” (1992: 32). This is 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.” He describes “eliminationism,” one major strand of the effort “to make philosophy scientific,” as “the view that mentalistic talk and mental entities would eventually lose their place in our attempts to describe and explain the world” (Burge 1992: 33), perhaps wrong, but surely an important thesis. That is, however, less than obvious.

Consider Nagel’s notions “capable of description by physical science” and “described by physics.” What do they mean? He offers the example

of liquidity, with its “transparent” relation to the behavior of molecules. The relation can’t be all that transparent. A century ago molecules were regarded by leading physicists as convenient fictions, and states of matter, as later learned, were not “capable of description” by then-existing physics. True, a branch of science not then unified with physics could provide much illumination in terms of its own theoretical constructs, as of much else; but the same is true today of some of the domain of the mental (in my sense). Why are these accounts less “physical” than chemistry was a century ago? Or less physical than Newton’s occult forces, and on to today’s arcane and counterintuitive theoretical posits? Perhaps the naturalistic accounts of mental phenomena will some day be unified with physics, which may again have to be revised, in which case the relations will also become “transparent.”

As for the thesis of eliminativism in Burge’s (again standard) formulation, we may ask why it is of any significance. Replace “mental” by “physical” in the thesis. Uncontroversially, “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 that enter into our common discourse and thinking. Why should we expect anything different of “mentalist talk and mental entities”? Suppose I say, “the rock dropped from the skies, rolled down the hill, and hit the ground.” The statement cannot be translated into the theories that have been developed to describe and explain the world, nor is there any interesting weaker relation; the terms belong to different intellectual universes. But no one takes this to constitute a body-body problem. Nor do the natural sciences aspire to distinguish this description from the statement that the rock fell down a crevice, which could be the same event viewed from a different perspective (with the hill not distinguished from surrounding terrain). Methodological naturalists do not expect to find counterparts to such informal statements as these within the explanatory theories they self-consciously devise; nor of “John took his umbrella because he thought it was going to rain,” or “John is in pain,” or “John speaks English” – though they hope, in all cases, that naturalistic inquiry might yield understanding and insight in the domains opened to inquiry by discourse reflecting common-sense perspectives.

Similar questions arise quite broadly. Take Donald Davidson’s “anomalousism of the mental,” the view that, while there are causal relations between mental and physical events, there are no psychophysical laws that connect them in an appropriate explanatory scheme. As Davidson puts it, one should not compare truisms about what people generally will do under certain conditions “with a law that says how fast

a body will fall in a vacuum,” because “in the latter case, but not the former, we can tell in advance whether the condition holds, and we know what allowance to make if it doesn’t,” (Davidson 1980: 233) a position on the mind–body problem that Burge describes as “profound but controversial” though inadequately clarified. (For a sympathetic discussion, see Evnine 1991.) The argument does not seem entirely compelling. For the same reason we should also not compare truisms about balls rolling down hills or a storm brewing in the West with the law of falling bodies, but we are not concerned about the lack of “physico-physical laws” connecting ordinary discourse about events in the world and explanatory theories of nature. It is argued that “folk psychology” is different from, say, “folk mechanics” or “folk chemistry” because of its *a priori* character and intimate relation to notions of rationality, reasons, intentions, first-person perspective, and so on. The domains are surely different, but it is unclear that they differ in “anomalism” in the sense of the discussion. Insofar as scientific inquiry might undermine one’s conviction that the Sun is setting or that objects are impenetrable (while leaving such convictions in place in other parts of life), it seems that it might in principle have similar effects on one’s convictions about the nature of beliefs (say, with regard to the role of rationality). Much of what people believe about beliefs is *a posteriori* (consider the debates about holism and innateness) and we have *a priori* beliefs about balls rolling down hills and storms brewing. Folk mechanics (etc.) seems no more susceptible than folk psychology to the formulation of bridge laws. As Davidson argues, mental event tokens are not tokens of physical event types (under informal description). The same is true of physical event tokens and physical objects, as common-sense construes them; only by fantastic accident will human language have natural kind terms, if natural kinds are the kinds of nature.<sup>2</sup>

To change terminology slightly, let us speak of “events mentalistically described” (“m-events”) and “events physicalistically described” (“p-events”), referring to accounts in ordinary language, reserving the terms *mental*, *chemical*, *optical*, etc. for events postulated by naturalistic inquiry in the mental, chemical, optical, etc. domains – all of these being “physical events,” a redundant term for events; same for objects, and so on. Then we expect to find causal relations between m-events and physical events, but no laws connecting them within explanatory science; the same is true of p-events. Beliefs, desires, perceptions, rocks rolling towards the ground, storms brewing, etc. are not subject to scientific laws, nor are there bridge laws connecting them to the sciences. Uncontroversially, science does not try to capture the content of ordinary discourse, let alone more creative acts of imagination. Paraphrasing



Nagel, we cannot “find a place in the world” of physics for physical phenomena, as we describe them in physicalistic talk (p-phenomena), so it is not surprising that the same is true of m-phenomena as captured in mentalistic talk.

Perhaps one should stress again that the reach of naturalistic inquiry may be quite limited, not approaching questions of serious human concern, however far-reaching its intellectual interest may prove to be. That is surely the present condition, and might so remain. Eliminativism, Nagel comments caustically, dismisses the “primitive theory” that was “the province of such simple folk as Flaubert, Proust, and Henry James.” Eliminativism does not seem to me a coherent position, but naturalism will hardly seek to annex this province, any more than it incorporates such trivial matters as rocks rolling down hills and storms brewing; on the contrary, it frees the explorer from irrelevant demands (see note 1).

Note that the truth of normal physicalistic talk and the status of the entities it postulates are not in question here. These are different topics. Nor is any question raised about the study of common-sense concepts as a branch of naturalistic inquiry (ethnoscience). It is interesting to learn how notions of language appear in the culture of the Navajo, (for an enlightening account, see Witherspoon 1977) or on the streets of New York, or even in the more self-consciously contrived culture of academic philosophy. The same is true of notions of physical objects and interaction, space, life and its origins, and so on. But such endeavors have to be taken seriously; they are not casual pursuits, and are not to be confused with naturalistic inquiry into the nature of what folk science addresses in its own ways, using possibly different faculties of mind. Ethnoscience is a branch of science that studies humans, seeking to understand their modes of interpretation of the world, the diversity of these systems, and their origins. Separate branches of science study the nature of what humans are sorting out and interpreting in their peculiar ways, whether the phenomena are optical, electrical, mechanical, or mental. Meanwhile, we continue to employ our concepts, sometimes choosing reflectively to refine and modify them, in trying to deal with the problems of ordinary life. These are distinct pursuits.

Ethnoscience asks how people interpret and evaluate what they find around them. It is concerned with accounts of objects striving to reach their natural place and of the motion of the heavenly bodies against the fixed stars; of the basic substances earth, air, fire, and water and how they combine to yield the phenomena of nature; of vital forces that guide biological development and differentiation; of beliefs, desires, fears, and other elements that enter into accounts of purposive action; and so on. It is not a trivial empirical claim that in some cultural tradition people



interpret motion in terms of contact; or, along Davidsonian lines, that they attribute beliefs and desires in terms of criteria of rationality and normativity with a holistic perspective, in their efforts to evaluate actions. These are strong claims, requiring evidence. It might turn out that beliefs and desires are attributed to creatures (perhaps humans) on entirely different grounds, perhaps as a reflection of instinctive modes of interpretation determined by innate endowment (common-sense), and that such attributions are systematically made even when the agents observed are considered to be acting in utterly irrational ways, or driven by instinct in contexts in which the question of rationality does not arise.

Whatever the ethnoscientist may discover about the nature of an “intentional stance” in Daniel Dennett’s sense, two further directions for scientific inquiry open up. One is about people: what are the origins of their modes of understanding; specifically, what role does innate endowment play in developing a cosmology, or judging that another person is reaching for a book or reading one, or hurrying to catch the bus. A second direction considers the topics that people are attempting to understand in the instinctively grounded and culturally shaped ways of the folk sciences. What is the truth about cosmology, the formation of continents, the diversity of insects, planning one’s actions, and so on. The answers, insofar as they are accessible to human intelligence, will be framed in terms appropriate to the problems at hand, with little concern for the intellectual apparatus of the folk sciences, and no expectation that constructs and principles that are developed will receive direct expression in terms of more “fundamental” branches of science, even if the unification problem has been solved. The end result may be to explain why folk-scientific interpretations more or less work, whether they are concerned with planets and flowers, or with a master chess player or a child building a tower with blocks (see Burge 1992; for some comments on attribution of mental states, in this context, see Chomsky 1969).

Returning to the critique of materialism – say, along Nagel’s lines – it seems to face several problems. The presupposed concepts “physical” or “material” have no clear sense; nor will “mental,” unless some sense can be given to the notion of *potential* consciousness and, even then, it is unclear what the interest of this particular category would be, as distinct from many others. It is not the business of the sciences to express the content of ordinary discourse about anything, physical or mental. There seems to be no coherent doctrine of materialism and metaphysical naturalism, no issue of eliminativism, no mind–body problem.

Problems mount when we look at how specific empirical questions are addressed. Nagel considers one: the proposal that there is a “Language Acquisition Device [LAD], which allows a child to learn the grammar of a language on the basis of the samples of speech it encounters” (1993: 109). He considers this a reputable part of science, right or wrong. But it is incorrect, he argues, to describe LAD as a “psychological mechanism,” as I do: it should be seen as “simply a physical mechanism – for it is incapable of giving rise to subjective conscious thought whose content consists of those rules themselves” (p. 109). Putting aside this conception of “the essence of mind” and the accuracy of the description of LAD (which I would not quite give this way), note that Nagel’s assertion appears to be an empirical one about the “capability” of some physical system. Again, we have the crucial matter of “potential consciousness,” now presented as an empirical hypothesis. We return to that.

What would be the reaction to a theory of LAD (of UG) by an avowed “eliminative materialist,” say Quine, whom Burge identifies as the originator of the doctrine? Quine puts forth the “naturalistic thesis” that “The world is as natural science says it is, insofar as natural science is right” (Quine 1992: 9); but that is not informative until we are told what “natural science” is. I suggested several possible answers, but Quine seems to have something else in mind. He takes natural science to be “theories of quarks and the like.” What is “like enough” to be part of science? Neurons are evidently allowed, along with certain psychological processes: thus language, Quine asserts, “is linked to our neural input by neural mechanisms of association or conditioning.” The empirical evidence is overwhelming that association and conditioning have little to do with language acquisition or use, but that seems not to matter; one wonders why. Whatever the answer, we find examples of what Quine favors (quarks, neural inputs, conditioning) and disfavors (the devices of LAD, that is, the operative mechanisms, so far as is known). But we are offered no reasons for the decisions, or more than a few examples to suggest their scope.

The “naturalistic thesis” proposed reveals the same arbitrariness in other domains. Thus Quine reiterates here the view he has often expounded that “reification of bodies comes in stages in one’s acquisition of language,” the “last stage” being recognition of identity over time. If that is an empirical hypothesis, one wants to know how it can be put forth with such confidence. It is surely not obvious, or even particularly plausible. We need not keep to anecdotal evidence; infant studies of past years provide considerable reason to believe that such “reification” appears in the first few months of life, long before any manifestation of

language. (For general review, see Spelke 1990; on more recent work, see Baillargeon 1993; see also note 3 of this chapter.)

Since the theories of LAD to which Nagel refers dismiss the dogmas about association and conditioning, and postulate mechanisms that are not (at least currently, maybe never) expressible in terms of quarks or neurons, they presumably do not fall within science, in Quine's sense. This is much like chemistry a century ago, or celestial mechanics at the time of Newton, by similar reasoning. Perhaps the empirical investigation of "reification" also fails Quine's criteria, for the same reason.<sup>3</sup> We seem to be faced with an extreme example of methodological dualism, over and above the obscure character of the notions "materialism" and "eliminativism."

### Access to consciousness

Let us turn now to the characterization of the mental in terms of access to consciousness, yielding the mind–body distinction, many hold. Adopting this characterization, Nagel concludes that LAD (and the state attained, an I-language, henceforth language) is only a physical mechanism, not a psychological mechanism, "for it is incapable of giving rise to subjective conscious thought whose content consists of those rules themselves" (1993: 109). Suppose that one option of variation among languages has to do with left–right orientation, English being syntactically "left-headed" ("see – the book," "in – the room," etc.) and Japanese "right-headed" (the mirror image, throughout). However, Johnny is not aware, and cannot tell us, that he is setting the "head parameter" as left–right on the basis of the evidence "see the book," etc., though perhaps that is exactly what is happening. Similarly, Mary has no conscious awareness that she is using principle (C) of binding theory when she interprets example (1) differently from example (2), excluding the option of referential dependence of *he* on *Bill* in example (1) but allowing it in example (2). Thus she does not interpret example (1) as (1') but may interpret example (2) as (2') (*he* = *Bill* in both cases):

- (1) He thinks Bill is a nice guy.
- (2) The woman he married thinks Bill is a nice guy.
- (1') Bill thinks he is a nice guy.
- (2') The woman Bill married thinks he is a nice guy.

Furthermore, this lack of awareness reaches to "potential consciousness," a notion yet to be clarified. Perhaps it means that no creature with Mary's language faculty, with these "physical mechanisms," can

have the consciousness Mary lacks, an important empirical truth. Consequently, the theories of LAD and of the language do not cross the body–mind divide; they are not about the mind, about psychological mechanisms.

To take an example from a different domain, Mary is not consciously aware that she is using a *rigidity principle* that interprets visual presentations as a rigid object in motion when she sees what she takes to be a cube rotating in space. And three-month-old Johnny cannot tell us, and may not be aware, of the beliefs about object constancy (“reification”) and trajectory that lead him to expect an object to appear in a particular form, time span, and place after passing behind a barrier (see Spelke 1990; Baillargeon 1993). Accordingly we cannot speak of the states and properties attributed to Mary and Johnny as psychological mechanisms of vision – at least, if potential consciousness is also lacking in these cases.

A similar idea is presented by Michael Dummett, though with different terminology. He regards the theories of LAD and the language attained as “psychological hypotheses,” though neither offers a “philosophical explanation,” because they do not tell us “the form in which [the body of knowledge] is delivered”; conscious awareness would, however, carry us past that divide (Dummett 1991: 97). Presumably the same would hold with regard to object constancy and the like. Here the distinction is not mind–body, but science–philosophy. For the sciences, the theories (accuracy aside) tell us everything relevant about the form in which the body of knowledge is delivered; however for the theory of meaning, (and, presumably, language and thought generally, and perhaps vision, reification, etc.), some additional kind of explanation is required, a “philosophical explanation,” that goes beyond science.

In both cases, we have a crucial distinction – perhaps a metaphysical distinction – based on access to consciousness.

Nagel’s account follows Searle’s in the book he is reviewing (see Burge 1992). We can trace the argument in its contemporary form back to Quine’s influential distinction between “fitting” and “guiding.” Quine objects to a traditional doctrine (reinterpreted within contemporary linguistics) that speakers are “guided” by a perhaps unconscious “notion of structure” in forming and interpreting newly created “free expressions” (Jespersen 1924: 19). This is an “enigmatic doctrine,” Quine holds, perhaps pure “folly” (Quine 1972: 447). We may speak of guiding only when rules are consciously applied to “cause” behavior; otherwise we may only say that behavior “fits” or “obeys” some system of rules, just as a planet obeys the law of falling bodies, and we must not attribute “psychological reality” to some particular conception of the nature of the organism that “obeys” the rules.

Once again, Quine adopts an extreme form of dualism. In the case of falling bodies, we are permitted – indeed, enjoined – to attribute “physical reality” to a particular conception of their nature and to the postulated principles. Plainly, we cannot account for the state attained by the language faculty and the ways it enters into behavior simply on the assumption that the brain has mass and obeys the law of falling bodies. More structure is needed. A naturalistic approach would proceed exactly as in the case of planets and ants; in this case, seeking a theory of initial and attained state, the relation between them, and the relation of the attained state to performance and judgments, attributing “reality” to whatever is postulated in the best theory we can devise. The level of understanding is far less in the vastly more intricate case of complex organisms, but that is not pertinent here.

A doctrinal divide is held to separate the cases: what is required in one case (falling bodies) is barred in the other (humans “above the neck”). Again, consciousness makes the difference, along with “causation of behavior,” a notion with its own nontrivial problems. We have little reason to believe that normal behavior *is* caused, at least in any known sense of that term, nor would a methodological naturalist dogmatically assume otherwise.

Quine’s reasoning would seem to apply in the same way to the visual example. Johnny and Mary are not “guided” by principles of rigidity, object constancy, and so on. Their behavior only “fits” these principles, as Mars satisfies the law of falling bodies. A theory of states of the brain that incorporates such principles to account for Mary’s and Johnny’s behavior, however well it might meet naturalistic standards, is methodologically defective; at best enigmatic, at worst folly. (As mentioned, Quine’s view on this matter is difficult to determine. See note 3.)

These ideas appear in many other variants. They are not easy to assess. Thus, no plausible reason is given for the strictures, no indication that they are more than terminological demands of no particular interest. The most developed version is the one that Nagel adopts from Searle. Let us briefly look at that.

The unexplained dualism of Quine’s distinction does not seem to have elicited much concern, but many see the consequences of the specific formulation as counterintuitive. Take the phenomenon of blindsight: Alice, who has sustained cerebral damage, distinguishes reliably between visual presentations (say, a drawing of a house on fire and of one that is not), but insists that they are identical, lacking any awareness of what enters into her differential behavior. In Quine’s terms, we cannot speak of guiding here, only fitting (so it seems; see Quine

1992: 9; note 3). In other versions, we cannot attribute to Alice “mental representations,” though we could for John, who is aware of and reports the difference, as Alice did before her injury. In Alice’s case we have only “physical mechanisms,” in John’s “psychological mechanisms”; or, in a different usage, for Alice we have only a “psychological hypothesis,” not a “philosophical explanation,” as we do for John. None of these seem attractive consequences.

Searle hopes to avoid them by introducing the notion of access to consciousness *in principle* – what Nagel, in his review, calls the *potential* for consciousness.<sup>4</sup> Searle’s “Connection Principle” (CP) requires accessibility in principle for attribution of mental states and processes. In the case of blindsight, Searle holds that Alice has access in principle to the representation, or the rule, or whatever. Blindsight is a case of mere “blockage,” not “inaccessibility in principle,” so we may speak of mental processes in Alice’s case, as in John’s. The conclusion will have substance when the term “in principle” is explained.

Suppose that Jane is identical to Alice (in relevant respects, a qualification henceforth omitted), except for her history: her neural condition was not the result of post-natal injury but of an injury at conception, which led to the condition. Presumably she too has “access in principle”; CP still holds (otherwise, the whole exercise was pointless; the time of the injury can hardly be relevant). Suppose that this injury at conception affected the genes in such a way as to yield blindsight; again, presumably, CP holds or the results are no less counterintuitive. Proceeding, suppose that Susan is identical to Jane except that the genetic change was due to a mutation, so that she is identical to Jane in genetic constitution, though she did not suffer blindsight through injury, as Alice and Jane did. Again, CP must hold, or the exercise was pointless. Susan, then, suffers only “blockage.” Suppose that Susan’s genetic property is transmitted, leading finally to a new subspecies. We now have the John-species and the Susan-species, exactly alike in their perceptual mechanisms. Members of the Susan-species are unaware of and cannot report the mental representations and rules that guide them. But the two subspecies are otherwise indistinguishable; and there is even some cross-species identity of visual mechanisms, as in the case of Alice and Jane after injury. Since CP holds of Susan, it presumably holds of the Susan-species; otherwise, again, we have completely pointless terminological stipulations.

Let’s now take the language case. Suppose we discover that our evolutionary history matches that of the Susan-species. That is, our forebears were actually a John-species, fully aware of how they set the head parameter, determined referential dependence, and so on, and

able to describe it all clearly to Martian scientists observing them. But a mutation took place (or perhaps an injury causing a genetic change, as in the case of Jane) and propagated, leading to us, a Susan-species, deprived of this ability. Suppose we even discover that we just haven't tested the right informants yet. The two subspecies are intermingled, and behave exactly alike; short of inquiring into awareness, none of us, and no scientist, can find any difference among the members. CP held for the earlier John-species, and for its remnants among us; hence for us as well, unless we choose to make terminological decisions which, as before, reveal the whole endeavor to be pointless.

But this is, completely the wrong result. The whole point of the exercise was to show that the naturalistic inquiry into language and mind does not yield "psychological reality," or "psychological mechanisms," or "philosophical explanations," or "mental representations," or "guiding" by rules. Crucially, the CP must determine that we have no access to the mechanisms and their operation *in principle*. We do not suffer mere "blockage"; rather, the mechanisms of our brain are "incapable of giving rise to subjective conscious thought whose content consists of those rules themselves" (Nagel 1993: 109), because all this lies beyond *potential* consciousness.

To save the story, it seems that we must insist that the John-species cannot exist in the case of language (though it can, and does, in the case of blindsight, namely humans): it is impossible for there to exist an organism exactly like us except that it is fully conscious of the content of the rules that it is following as it learns (and uses) a language. That, at last, looks like an empirical hypothesis, not a terminological stipulation. On what basis do we assert it? Or if the claim is not empirical but conceptual, what are the grounds for it? And whether we accept it or not – whether as an empirical or conceptual thesis – what possible interest does it have? How does it differ from some pronouncement about "the essence of the chemical" (electrical, optical, etc.)?

Similar questions arise in the case of object perception discussed earlier, and difficulties can be elaborated, leading to still further paradox. None of these questions arise in naturalistic inquiry, which has no place for such notions as "access in principle" or "potentially conscious" or CP, no notion of "philosophical explanation" beyond explanation, no privileged categories of evidence (like awareness, or "psychological" versus "linguistic" evidence), no mind–body distinction, no methodological (or other) dualism.

The effort to maintain these dualisms is reminiscent of attempts to salvage the idea that knowledge is a kind of ability, in the face of the fact that ability can improve or decline – or even be completely lost

– while knowledge remains unchanged, as illustrated, for example, by loss of ability to speak (swim, etc.) after injury and recovery without relevant input as effects of the injury recede. The natural conclusion is that knowledge (how . . . , that . . . , or whatever) has a crucial cognitive element, and ability to use knowledge is not to be confused with knowledge. To avoid the conclusion, a new technical concept with the properties of knowledge is constructed – called “ability” – but distinct from the ordinary concept, a move that is particularly odd when undertaken in alleged defense of a Wittgensteinian point of view. (See note 4 for references and discussion.)

### Further varieties of dualism

Much of the discussion of rule-following in the literature takes as a model arithmetical or traffic rules, or those given in grammar books, or others with a normative character. A crucial feature of rule-following, then, is that error must be possible in the sense of violation of the norm. Whatever the interest of this discussion, it is not to the point here. Rules of language – for example, the principles of UG, or those that guide Mary’s judgments about examples (1) and (2) above (see p. 93) – are not normative in this sense. Mary’s judgments and other behavior can be “in error,” for any number of reasons; for example, inattention or parsing difficulty (as in “garden path” sentences, or expressions that overwhelm perceptual capacities). Mary can also decide to violate her rules, perhaps for quite good reasons, say for literary effect. Judgments and behavior may also be inconsistent with norms in many ways: norms stipulated in various authoritarian structures, common practice in communities of the endlessly varying sort that individuals may be associated with, by choice or external pressure, and so on. Numerous questions arise of fact, policy, etc., but there seem to be none of principle, apart from questions that reduce to skeptical arguments of no special interest in this connection. (For further discussion, see Chomsky 1986.)

Should we speak of “following rules” in the case of Mary’s linguistic judgments and behavior? The question is not very interesting, for reasons already mentioned; no one expects common discourse to survive the transition to explanatory theory. However, for the record, to speak of Mary as following rules in this case would be closer to common usage than the standard philosophical convention that requires a link to consciousness. In fact, it keeps fairly close to common usage except in one respect. We typically use the term “rule-following” in the case of *deviation* from community norms, not observance of them, as in the



technical usage of philosophical discourse. Thus if Johnny says “I brang my lunch home,” normal usage would be that he is following the rule for “sing,” etc. – mistakenly, in that authority figures or some other standards call for “brought.” Similarly if he uses “puppy” to refer to kittens, following the rule that small household pets are puppies. Someone who is attentive might make comparable comments about the rules of pronunciation he follows. If all adults were to die and Johnny and his cohorts to survive, they would continue to follow their private and individual rules, except that now these would be rules of a perfectly normal human language which differs from standard English in these (and other) respects. In that case, however, we would not normally say that Johnny is following a rule, because the term is rarely used for observance of norms and standards. Thus, only linguists would say that Mary is following Principle (C) of binding theory in examples (1) and (2), or following the intricate and complex rules of referring to objects when she talks about her house.

When we attribute rule-following in the normal way – say, to Johnny in the case above – we do not mean to suggest that the rule-followers are (or could be) aware of following the rules or choosing to do so. Those who speak of “the fact that linguistic meaning involves deliberate rule-following” are using the term “rule-following” in a technical sense of philosophical discourse, not in the conventional way (Baldwin 1993: 187, citing P. Pettit). I think the same is true of other terms of philosophical discourse, including “knowledge,” “content,” and “reference,” among others. For some discussion, see publications cited earlier, and Chapter 2 of this volume.

Within the naturalistic theory of (I-)language – internalist and individualist – conclusions can be drawn about what one ought to do, but only in uninteresting hypothetical imperatives (if you want to rhyme something with “tower” or refer to daffodils, use “flower” not “book”). Such normativity, a regular consequence of knowledge, abounds in a naturalistic setting, but not the kind that arises when we ask whether Jones should change his usage of “arthritis” to conform to that of the doctor, a question of a very different kind, with no definite answer apart from specification of one or another region in a highly intricate space of human interests and concerns.

A related matter is the notion of language as a “community property” of some kind, as when we say that Hans and Maria speak German even though they cannot understand one another, and Hans does not speak Dutch though he understands quite well the Dutch spoken right across the border. Or when we say that Pierre and his son Jean, monolingual speakers of French who have moved to New York, are learning English,

which Jean will succeed in doing though Pierre only partially so. Or that Johnny, with his “mistakes” about “brang,” “puppy,” and the pronunciation of his name, speaks no language at all (an odd gap in normal usage), though he will speak English some day and has “partial knowledge” of it today, and his current I-language would be a normal language if it perpetuates as described. A vast range of such usages are not problematic in ordinary life, but are of little interest for the effort to understand what language is and how it is used. It is not a matter of idealization; there are no sensible idealizations, any more than we reify *areas* in clarifying what is meant by the statement that John lives near Mary but far from Bill. Sometimes these usages are codified in “national languages,” sometimes even imposed by force. Attempts to relate notions of “common language” to cultures simply make matters worse. A person will typically be part of many communities and cultures, with only weak correlations among the forms of association. Jones may participate in a common culture – with shared values, beliefs, understanding, etc. – with a monolingual speaker of some language he knows not a word of, perhaps to a greater extent than with his identical twin, with whom he grew up and whose speech is virtually indistinguishable from his own. None of this has anything to do with successful communication. We need not assume shared pronunciations or meanings to account for this, any more than we assume shared shapes to account for people who look alike.

Again, one may describe the innumerable situations that arise, and study of them is legitimate and useful. If pursued seriously, such study presupposes what is learned from naturalistic inquiry into the language faculty. However, attempts to base theories of pronunciation or meaning (with common pronunciations and common meanings) on alleged community properties can only lead to confusion. Such attempts again illustrate the kind of dualism that would never be taken seriously beyond the domain of the mental.

Another form of dualism that has arisen in the discussion of language acquisition is illustrated by a curious debate on “innatism” or “the innateness hypothesis.” The debate is one-sided: no one defends the hypothesis, including those to whom it is attributed (me, in particular). The reason is that there is no such hypothesis. There are certain proposals about the initial state of the language faculty (LAD, UG). These are not questioned by the critics. Rather, they regard the enterprise as somehow faulty, apparently on some dualist assumption. Similar questions are not raised when proposals are made about other aspects of growth, and no reason has been given as to why they are appropriate here. Alternative theses of a very general nature have been put forth: for

example, that “general learning mechanisms” suffice, with no need to assume specific properties of the language faculty. Such theses cannot be discussed until we are told what these mechanisms are. Specific proposals that have been made are hardly worth considering on naturalistic grounds, so they must be motivated by some other demands, dualist in nature.

Quine’s behaviorism is a variant of this form of dualism.<sup>5</sup> He argues that “the behaviorist approach is mandatory” (Quine 1990: 37) for the study of language because, in acquiring language, “we depend strictly on overt behavior in observable situations” (p. 38). By similar argument, the nutritionist approach is mandatory in embryology because, in the passage from embryo to mature state, the organism depends strictly on nutrition provided from outside; just as linguists must be behaviorists, so biologists must be nutritionists, restricting themselves to observation of nutritional inputs. The fallacy in the latter argument is apparent; the same fallacy undermines the former. Only radical dualist assumptions allow the matter even to be discussed. Perhaps the actual study of language is conceptually flawed but, in order to establish this, it does not suffice to demand that the linguist abandon naturalistic inquiry – as Quine and his followers do – and adopt stipulations that are arbitrary apart from their historical antecedents, plainly irrelevant.

Closely related is Quine’s radical translation paradigm. In the naturalistic study of interaction among organisms (cells, insects, birds, dolphins, . . .), we try to discover what internal states make the interaction possible, yielding the interpretations given to signals. In the study of human language, that path is interdicted. The study of interaction must keep within stipulated bounds: the investigating scientist is permitted to register noises in a specific way, to pick out some features of the situation, to test assent or dissent to the inquiry “Is this an X?,” and to carry out elementary induction, but nothing more. Various hints are given as to the features admissible, the choice of X, etc. Quine alleges further that this is also the epistemic situation of the child acquiring language and the person in a communication interchange. The three cases are entirely different in character: the child comes equipped with the initial state of the language faculty (LAD, UG); the person in a communication interchange, with the properties of the attained state; the linguist, with the science-forming faculty and the result of earlier inquiry into language. It is not, however, important to sort this out, because there is a more fundamental problem: the radical dualism of the entire approach. Nothing remotely similar would be tolerated in the study of other organisms, or aspects of humans that do not fall under the traditional descriptive category of the “mental.”

From this paradigm, widely adopted and discussed, far-reaching conclusions are drawn about language and thought. It appears to be a pointless intellectual exercise if intended to shed light on the nature of communication, acquisition, or the study of language and thought. At least, no satisfactory justification has been offered for it, to my knowledge, nor any explanation of why the approach should be adopted (or even considered) in this unique case. If the goal is to sharpen understanding of the concepts of belief, intention, meaning, and the like, the criteria for evaluation are more obscure but it is hard to see why the specific stipulated conditions should be privileged in this conceptual inquiry.

The paradigm underlies other dualist moves. Adapting it to his own concerns, Davidson argues that the goal of the descriptive study of meaning is to construct a theory that “is a model of the interpreter’s linguistic competence,” but that it “does not add anything to this thesis to say that if the theory does correctly describe the competence of an interpreter, some mechanisms in the interpreter must correspond to the theory” (Davidson 1986b: 438). Like Quine, he stipulates what is to count as relevant evidence: “what is open to observation is the use of sentences in context,” nothing more. Theories may introduce “reference and related semantic notions,” but there “can be no question about the correctness of these theoretical concepts beyond the question of whether they yield a satisfactory account of the use of sentences” (Davidson 1990: 300). Similar positions have been developed by Dummett and others (see Davidson 1986b; 1990a; on Dummett’s version, see Chomsky 1986).

Again, comparable ideas would not be taken seriously in the study of other systems. Only if we keep to the radical translation paradigm or some other arbitrary constraint is evidence restricted to the use of sentences by the speaker (or some selected community). Approaching the topic as in the sciences, we will look for all sorts of evidence. For example, evidence from Japanese will be used (and commonly is used) for the study of English; quite rationally, on the well-supported empirical assumption that the languages are modifications of the same initial state. Similarly, evidence can be found from studies of language acquisition and perception, aphasia, sign language, electrical activity of the brain, and who knows what else. Furthermore, it adds a great deal to postulate mechanisms in the interpreter that “correspond to the theory,” since it is precisely that move that subjects the theory to a wide variety of evidence beyond the stipulations of radical translation. Davidson’s injunction simply bars naturalistic inquiry into the nature of the interpreter. Efforts to verify and improve the postulated account are declared illegitimate, or perhaps irrelevant for some reason. The same holds of many other variants.

In his historical reconstruction of the origins of the “Theory–Theory,” Stephen Stich observes that with “the decline of Cartesian dualism, philosophers began looking for a way to locate the mental *within* the physical, identifying mental events with some category of events in the physical world” (Stich 1983: 14). That quest could have taken two directions, he observes: an attempt “to define mental vocabulary in *neurological* terms” (p. 14), or an analysis of mental concepts in terms of behavior, leading to philosophical behaviorism. The latter prevailed, he argues. What has just been reviewed is one highly influential strand, with no redeeming features, as far as I can see. The other direction has also been pursued, but also tainted by unjustified dualism.

Before turning to that, a few comments on this way of framing the issues. First, the reasons for the collapse of Cartesian dualism are somewhat misconstrued: as noted, it was the theory of body that was refuted, leaving no intelligible mind–body problem, no notion of “physical,” etc. In this realm, we have only the naturalistic approach: to construct explanatory theory in whatever terms are appropriate, and to face the unification problem. Second, it is, for the moment, only a hope that “neurological terms” are relevant for the unification problem. Finally, there is no reason to try to define the “mental vocabulary” of ordinary discourse in a naturalistic framework, just as no one contemplates that for “physical vocabulary,” at least in the modern period. Stich reaches a similar conclusion, but it is not clear why it even requires argument, dualist prejudice aside.

Naturalistic inquiry into the mind yields theories about the brain, its states and properties: UG, for example. No one knows how to begin to relate these theories to properties of atoms, cells, neurons, or other known structures of the brain. The disparity between theories of the mind and what has been learned about neurophysiology “creates a crisis for those who believe that the nervous system is precise and ‘hardwired’ like a computer,” biologist Gerald Edelman (1992: 27f.) concludes; and for connectionist and neural net theories as well. The varied individual histories of the nervous system and the “enormous individual structural variation” of brains provide “the coup de grace (actually multiple coups!)” (Edelman 1992: postscript) to attempts to construct computational or neural net theories of the mind. Apparently Edelman takes this to be true no matter how successful such studies might be, now or ever, by the standards of science (explanation, insight, etc.).

By similar logic, one could have argued not long ago that there is a terrible crisis for the study of matter and organisms in terms of colors, valence, the solid state, and a multitude of other properties; and earlier, for the investigation of electricity and magnetism, planetary and celestial

motion, etc. Virtually the whole of science was in crisis because of the huge gap between what had been learned about these topics and the principles of the mechanical philosophy (or even much more recent physics). The crisis Edelman perceives is real, but misplaced.

As for the “enormous variation” in structure of brains and experience, that tells us little. Not many years ago, languages appeared to differ from one another as radically as neural structures do to many a trained eye today, and were considered mere reflections of infinitely variable experience. Any complex system will appear to be a hopeless array of confusion before it comes to be understood, and its principles of organization and function discovered. Edelman argues that introducing considerations of meaning will somehow overcome the alleged problems of “formalist” approaches. These he seriously misunderstands – so his few comments indicate – but more important is the mistaken view of semantics. Simple semantic properties pose all the problems Edelman perceives in syntactic theories and constructions. They are rule-governed, sharply delineated, and fixed in relative independence of experience and known aspects of neural structure; hence they too induce the “crisis” caused by the gap between the apparent algorithmic, digital character of language and the observed variability and continuous flux of individual experience and neural structure. We face a typical problem of unification in the sciences, which may, as often in the past, require that the more “basic” science be fundamentally recast if it is to be integrated with successful explanatory theory at other levels.

Various remedies have been proposed to deal with the “crisis.” One is the proposal that “the mental is the neurophysiological at a higher level.” That could turn out to be true, but it is now a hypothesis about the neurophysiological, not a characterization of the mental; the shoe is on the wrong foot, in the light of what is at all understood. Another is the version of “eliminative materialism” that holds that we should concentrate on neurophysiology, which has all the merit of a proposal some time ago that chemistry should be abandoned in favor of the study of solid particles in motion, or that embryologists should follow the same course. There is a substantial literature asking what it would imply if neural-net (connectionist) models could account for the phenomena that have been explained in terms of computational-representational systems. Such discussion may appear to be naturalistic in temper, but that is hardly clear. Few biologists would be intrigued by the suggestion that unstructured systems with unknown properties might some day make it possible to account for development of organisms without appeal to complex constructions in terms of concentration of chemicals, the cell’s internal program, production of proteins, and so on.

In some domains – language in particular – successful theories are commonly of the computational–representational type, a fact that causes considerable uneasiness. To relieve it, computer models are often invoked to show that we have robust, hard-headed instances of the kind: psychology then studies software problems. That is a dubious move. Artifacts pose questions that do not arise in the case of natural objects. Whether some object is a key or a table or a computer depends on designer's intent, standard use, mode of interpretation, and so on. The same considerations arise when we ask whether the device is malfunctioning, following a rule, etc. There is no natural kind or normal case. These questions do not arise in the study of organic molecules, the wings of chickens, the language faculty, or other natural objects. The belief that there was a problem to resolve, beyond the normal ones, reflects an unwarranted dualism; the proposed cure is worse than the disease.

These remarks barely skim the surface of dualist elements in much of the most sophisticated and influential thinking about language and mind. These should either be justified or abandoned. The critique of naturalistic approaches also seems to me to be flawed. There is, I think, good reason to examine more closely doctrines that have been assumed too casually, and if they do not withstand such analysis, to ask why they seem so compelling.