THE MIT ENCYCLOPEDIA OF THE COGNITIVE SCIENCES

MIT 认知科学百科全书

主编

ROBERT A. WILSON
FRANK C. KEIL

The MIT Encyclopedia of the Cognitive Sciences

MIT 认知科学百科全书

Robert A. Wilson　Frank C. Keil　主编

The MIT Encyclopedia
of the Cognitive Sciences

EDITED BY

Robert A. Wilson and

Frank C. Keil

To the memory of Henry Bradford Stanton (a.k.a "Harry the hat"), 1921–1997, and to his wife Betty upon her retirement, after twenty-one years with Bradford Books. Harry and Betty were its cofounders and a major force in their own right in the flowering and cross-fertilization of the interdisciplinary cognitive sciences.

Preface

The *MIT Encyclopedia of the Cognitive Sciences* (*MITECS* to its friends) has been four years in the making from conception to publication. It consists of 471 concise articles, nearly all of which include useful lists of references and further readings, preceded by six longer introductory essays written by the volume's advisory editors. We see *MITECS* as being of use to students and scholars across the various disciplines that contribute to the cognitive sciences, including psychology, neuroscience, linguistics, philosophy, anthropology and the social sciences more generally, evolutionary biology, education, computer science, artificial intelligence, and ethology.

Although we prefer to let the volume speak largely for itself, it may help to provide some brief details about the aims and development of the project. One of the chief motivations for this undertaking was the sense that, despite a number of excellent works that overlapped with the ambit of cognitive science as it was traditionally conceived, there was no single work that adequately represented the full range of concepts, methods, and results derived and deployed in cognitive science over the last twenty-five years.

Second, each of the various cognitive sciences differs in its focus and orientation; in addition, these have changed over time and will continue to do so in the future. We see *MITECS* as aiming to represent the scope of this diversity, and as conveying a sense of both the history and future of the cognitive sciences.

Finally, we wanted, through discussions with authors and as a result of editorial review, to highlight links across the various cognitive sciences so that readers from one discipline might gain a greater insight into relevant work in other fields. *MITECS* represents far more than an alphabetic list of topics in the cognitive sciences; it captures a good deal of the structure of the whole enterprise at this point in time, the ways in which ideas are linked together across topics and disciplines, as well as the ways in which authors from very different disciplines converge and diverge in their approaches to very similar topics. As one looks through the encyclopedia as a whole, one takes a journey through a rich and multidimensional landscape of interconnected ideas. Categorization is rarely just that, especially in the sciences. Ideas and patterns are related to one another, and the grounds for categorizations are often embedded in complex theoretical and empirical patterns. *MITECS* illustrates the richness and intricacy of this process and the immense value of cognitive science approaches to many questions about the mind.

All three of the motivations for *MITECS* were instrumental in the internal organization of the project. The core of *MITECS* is the 471 articles themselves, which were assigned to one of six fields that constitute the foundation of the cognitive sciences. One or two advisory editors oversaw the articles in each of these fields and contributed the introductory essays. The fields and the corresponding advisory editors are

Philosophy (Robert A. Wilson)
Psychology (Keith J. Holyoak)
Neurosciences (Thomas D. Albright and Helen J. Neville)
Computational Intelligence (Michael I. Jordan and Stuart Russell)
Linguistics and Language (Gennaro Chierchia)
Culture, Cognition, and Evolution (Dan Sperber and Lawrence Hirschfeld)

These editors advised us regarding both the topics and authors for the articles and assisted in overseeing the review process for each. Considered collectively, the articles represent much of the diversity to be found in the corresponding fields and indicate much of what has been, is, and might be of value for those thinking about cognition from one or another interdisciplinary perspective.

Each introduction has two broad goals. The first is to provide a road map through *MITECS* to the articles in the corresponding section. Because of the arbitrariness of

assigning some articles to one section rather than another, and because of the interdisciplinary vision guiding the volume, the introductions mention not only the articles in the corresponding section but also others from overlapping fields. The second goal is to provide a perspective on the nature of the corresponding discipline or disciplines, particularly with respect to the cognitive sciences. Each introduction should stand as a useful overview of the field it represents. We also made it clear to the editors that their introductions did not have to be completely neutral and could clearly express their own unique perspectives. The result is a vibrant and engaging series of essays.

We have been fortunate in being able to enlist many of the world's leading authorities as authors of the articles. Our directions to contributors were to write articles that are both representative of their topic and accessible to advanced undergraduates and graduate students in the field. The review process involved assigning two reviewers to each article, one an expert from within the same field, the other an outsider from another field represented in *MITECS*; nearly all reviewers were themselves contributors to *MITECS*. In addition, every article was read by at least one of the general editors. Articles that did not seem quite right to either or both of us or to our reviewers were sometimes referred to the advisory editors. One might think that with such short articles (most being between 1,000 and 1,500 words in length), the multiple levels of review were unnecessary, but the selectivity that this brevity necessitated made such a review process all the more worthwhile. Relatedly, as more than one contributor noted in explaining his own tardiness: "This article would have been written sooner if it hadn't been so short!".

Of course the content of the articles will be the chief source of their value to the reader, but given the imposed conciseness, an important part of their value is the guide that their references and further readings provide to the relevant literature. In addition, each article contains cross-references, indicated in SMALL CAPITALS, to related articles and a short list of "see also" cross-references at the end of the article. Responsibility for these cross-references lies ultimately with one of us (RAW), though we are thankful to those authors who took the time to suggest cross-references for their own articles.

We envisioned that many scholars would use *MITECS* as a frequent, perhaps even daily, tool in their research and have designed the references, readings, and cross-references with that use in mind. The electronic version will allow users to download relevant references into their bibliography databases along with considerable cross-classification information to aid future searches. Both of us are surprised at the extent to which we have already come to rely on drafts of articles in *MITECS* for these purposes in our own scholarly pursuits.

In the long list of people to thank, we begin with the contributors themselves, from whom we have learned much, both from their articles and their reviews of the articles of others, and to whom readers owe their first debt. Without the expertise of the advisory editors there is little chance that we would have arrived at a comprehensive range of topics or managed to identify and recruit many of the authors who have contributed to *MITECS*. And without their willingness to take on the chore of responding to our whims and fancies over a three-year period, and to write the section introductions, *MITECS* would have fallen short of its goals. Thanks Tom, Gennaro, Larry, Keith, Mike, Helen, Stuart, and Dan. At The MIT Press, we thank Amy Brand for her leadership and persistence, her able assistants Ed Sprague and Ben Bruening for their tech-know-how and hard work, and Sandra Minkkinen for editorial oversight of the process.

Rob Wilson thanks his coterie of research assistants: Patricia Ambrose and Peter Piegaze while he was at Queen's University; and Aaron Sklar, Keith Krueger, and Peter Asaro since he has been at the University of Illinois. His work on *MITECS* was supported, in part, by SSHRC Individual Three-Year Grant #410-96-0497, and a UIUC Campus Research Board Grant. Frank Keil thanks Cornell University for internal funds that were used to help support this project.

Philosophy

Robert A. Wilson

The areas of philosophy that contribute to and draw on the cognitive sciences are various; they include the philosophy of mind, science, and language; formal and philosophical logic; and traditional metaphysics and epistemology. The most direct connections hold between the philosophy of mind and the cognitive sciences, and it is with classical issues in the philosophy of mind that I begin this introduction (section 1). I then briefly chart the move from the rise of materialism as the dominant response to one of these classic issues, the mind-body problem, to the idea of a science of the mind. I do so by discussing the early attempts by introspectionists and behaviorists to study the mind (section 2). Here I focus on several problems with a philosophical flavor that arise for these views, problems that continue to lurk backstage in the theater of contemporary cognitive science.

Between these early attempts at a science of the mind and today's efforts lie two general, influential philosophical traditions, ordinary language philosophy and logical positivism. In order to bring out, by contrast, what is distinctive about the contemporary naturalism integral to philosophical contributions to the cognitive sciences, I sketch the approach to the mind in these traditions (section 3). And before getting to contemporary naturalism itself I take a quick look at the philosophy of science, in light of the legacy of positivism (section 4).

In sections 5 through 7 I get, at last, to the mind in cognitive science proper. Section 5 discusses the conceptions of mind that have dominated the contemporary cognitive sciences, particularly that which forms part of what is sometimes called "classic" cognitive science and that of its connectionist rival. Sections 6 and 7 explore two specific clusters of topics that have been the focus of philosophical discussion of the mind over the last 20 years or so, folk psychology and mental content. The final sections gesture briefly at the interplay between the cognitive sciences and logic (section 8) and biology (section 9).

1   Three Classic Philosophical Issues About the Mind

i. The Mental-Physical Relation

The relation between the mental and the physical is the deepest and most recurrent classic philosophical topic in the philosophy of mind, one very much alive today. In due course, we will come to see why this topic is so persistent and pervasive in thinking about the mind. But to convey something of the topic's historical significance let us begin with a classic expression of the puzzling nature of the relation between the mental and the physical, the MIND-BODY PROBLEM.

This problem is most famously associated with RENÉ DESCARTES, the preeminent figure of philosophy and science in the first half of the seventeenth century. Descartes combined a thorough-going mechanistic theory of nature with a *dualistic* theory of the nature of human beings that is still, in general terms, the most widespread view held by ordinary people outside the hallowed halls of academia. Although nature, including that of the human body, is material and thus completely governed by basic principles of mechanics, human beings are special in that they are composed both of *material and nonmaterial or mental stuff, and so are not so governed.* In Descartes's own terms, people are essentially a combination of mental substances (minds) and material substances (bodies). This is Descartes's *dualism.* To put it in more commonsense terms, people have both a mind and a body.

Although dualism is often presented as a possible solution to the mind-body problem, a possible position that one might adopt in explaining how the mental and physical are related, it serves better as a way to bring out why there is a "problem" here at all. For if the mind is one type of thing, and the body is another, how do these two

types of things interact? To put it differently, if the mind really is a nonmaterial sub-stance, lacking physical properties such as spatial location and shape, how can it be both the cause of effects in the material world—like making bodies move—and itself be causally affected by that world—as when a thumb slammed with a hammer (bodily cause) causes one to feel pain (mental effect)? This problem of causation between mind and body has been thought to pose a largely unanswered problem for Cartesian dualism.

It would be a mistake, however, to assume that the mind-body problem in its most general form is simply a consequence of dualism. For the general question as to how the mental is related to the physical arises squarely for those convinced that some ver-sion of materialism or PHYSICALISM must be true of the mind. In fact, in the next sec-tion, I will suggest that one reason for the resilience and relevance of the mind-body problem has been the *rise* of materialism over the last fifty years.

Materialists hold that all that exists is material or physical in nature. Minds, then, are somehow or other composed of arrangements of physical stuff. There have been various ways in which the "somehow or other" has been cashed out by physicalists, but even the view that has come closest to being a consensus view among contempo-rary materialists—that the mind *supervenes* on the body—remains problematic. Even once one adopts materialism, the task of articulating the relationship between the mental and the physical remains, because even physical minds have special properties, like intentionality and consciousness, that require further explanation. Simply pro-claiming that the mind is not made out of distinctly mental substance, but is material like the rest of the world, does little to explain the features of the mind that seem to be distinctively if not uniquely features of physical minds.

ii. The Structure of the Mind and Knowledge

Another historically important cluster of topics in the philosophy of mind concerns what is in a mind. What, if anything, is distinctive of the mind, and how is the mind structured? Here I focus on two dimensions to this issue.

One dimension stems from the RATIONALISM VS. EMPIRICISM debate that reached a high point in the seventeenth and eighteenth centuries. Rationalism and empiricism are views of the nature of human knowledge. Broadly speaking, empiricists hold that all of our knowledge derives from our sensory, experiential, or empirical interaction with the world. Rationalists, by contrast, hold the negation of this, that there is some knowledge that does not derive from experience.

Since at least our paradigms of knowledge—of our immediate environments, of common physical objects, of scientific kinds—seem obviously to be based on sense experience, empiricism has significant intuitive appeal. Rationalism, by contrast, seems to require further motivation: minimally, a list of knowables that represent a prima facie challenge to the empiricist's global claim about the foundations of knowl-edge. Classic rationalists, such as Descartes, Leibniz, Spinoza, and perhaps more con-tentiously KANT, included knowledge of God, substance, and abstract ideas (such as that of a triangle, as opposed to ideas of particular triangles). Empiricists over the last three hundred years or so have either claimed that there was nothing to know in such cases, or sought to provide the corresponding empiricist account of how we could know such things from experience.

The different views of the sources of knowledge held by rationalists and empiricists have been accompanied by correspondingly different views of the mind, and it is not hard to see why. If one is an empiricist and so holds, roughly, that there is nothing in the mind that is not first in the senses, then there is a fairly literal sense in which *ideas,* found in the mind, are complexes that derive from *impressions* in the senses. This in turn suggests that the processes that constitute cognition are themselves elaborations of those that constitute perception, that is, that cognition and perception differ only in degree, not kind. The most commonly postulated mechanisms governing these pro-cesses are *association* and *similarity,* from Hume's laws of association to feature-extraction in contemporary connectionist networks. Thus, the mind tends to be viewed by empiricists as a *domain-general* device, in that the principles that govern its opera-

tion are constant across various types and levels of cognition, with the common empirical basis for all knowledge providing the basis for parsimony here.

By contrast, in denying that all knowledge derives from the senses, rationalists are faced with the question of what other sources there are for knowledge. The most natural candidate is the mind itself, and for this reason rationalism goes hand in hand with NATIVISM about both the source of human knowledge and the structure of the human mind. If some ideas are innate (and so do not need to be derived from experience), then it follows that the mind already has a relatively rich, inherent structure, one that in turn limits the malleability of the mind in light of experience. As mentioned, classic rationalists made the claim that certain ideas or CONCEPTS were innate, a claim occasionally made by contemporary nativists—most notably Jerry Fodor (1975) in his claim that *all* concepts are innate. However, contemporary nativism is more often expressed as the view that certain implicit knowledge that we have or principles that govern how the mind works—most notoriously, linguistic knowledge and principles—are innate, and so not learned. And because the types of knowledge that one can have may be endlessly heterogeneous, rationalists tend to view the mind as a *domain-specific* device, as one made up of systems whose governing principles are very different. It should thus be no surprise that the historical debate between rationalists and empiricists has been revisited in contemporary discussions of the INNATENESS OF LANGUAGE, the MODULARITY OF MIND, and CONNECTIONISM.

A second dimension to the issue of the structure of the mind concerns the place of CONSCIOUSNESS among mental phenomena. From WILLIAM JAMES's influential analysis of the phenomenology of the stream of consciousness in his *The Principles of Psychology* (1890) to the renaissance that consciousness has experienced in the last ten years (if publication frenzies are anything to go by), consciousness has been thought to be the most puzzling of mental phenomena. There is now almost universal agreement that conscious mental states are a part of the mind. But how large and how important a part? Consciousness has sometimes been thought to exhaust the mental, a view often attributed to Descartes. The idea here is that everything mental is, in some sense, conscious or available to consciousness. (A version of the latter of these ideas has been recently expressed in John Searle's [1992: 156] *connection principle*: "all unconscious intentional states are in principle accessible to consciousness.")

There are two challenges to the view that everything mental is conscious or even available to consciousness. The first is posed by the *unconscious*. SIGMUND FREUD's extension of our common-sense attributions of belief and desire, our folk psychology, to the realm of the unconscious played and continues to play a central role in PSYCHO-ANALYSIS. The second arises from the conception of cognition as information processing that has been and remains focal in contemporary cognitive science, because such information processing is mostly *not* available to consciousness. If cognition so conceived is mental, then most mental processing is not available to consciousness.

iii. The First- and Third-Person Perspectives

Occupying center stage with the mind-body problem in traditional philosophy of mind is the *problem of other minds,* a problem that, unlike the mind-body problem, has all but disappeared from philosophical contributions to the cognitive sciences. The problem is often stated in terms of a contrast between the relatively secure way in which I "directly" know about the existence of *my own* mental states, and the far more epistemically risky way in which I must infer the existence of the mental states of others. Thus, although I can know about my own mental states simply by introspection and self-directed reflection, because this way of finding out about mental states is peculiarly first-person, I need some other type of evidence to draw conclusions about the mental states of others. Naturally, an agent's behavior is a guide to what mental states he or she is in, but there seems to be an epistemic gap between this sort of evidence and the attribution of the corresponding mental states that does not exist in the case of self-ascription. Thus the problem of other minds is chiefly an *epistemological* problem, sometimes expressed as a form of skepticism about the justification that we have for attributing mental states to others.

There are two reasons for the waning attention to the problem of other minds *qua* *problem* that derive from recent philosophical thought sensitive to empirical work in the cognitive sciences. First, research on introspection and SELF-KNOWLEDGE has raised questions about how "direct" our knowledge of our own mental states and of the SELF is, and so called into question traditional conceptions of first-person knowledge of mentality. Second, explorations of the THEORY OF MIND, ANIMAL COMMUNICATION, and SOCIAL PLAY BEHAVIOR have begun to examine and assess the sorts of attribution of mental states that are actually justified in empirical studies, suggesting that third-person knowledge of mental states is not as limited as has been thought. Considered together, this research hints that the contrast between first- and third-person knowledge of the mental is not as stark as the problem of other minds seems to intimate.

Still, there is something distinctive about the first-person perspective, and it is in part as an acknowledgment of this, to return to an earlier point, that consciousness has become a hot topic in the cognitive sciences of the 1990s. For whatever else we say about consciousness, it seems tied ineliminably to the first-person perspective. It is a state or condition that has an irreducibly *subjective* component, something with an essence to be experienced, and which presupposes the existence of a subject of that experience. Whether this implies that there are QUALIA that resist complete characterization in materialist terms, or other limitations to a science of the mind, remain questions of debate.

See also ANIMAL COMMUNICATION; CONCEPTS; CONNECTIONISM, PHILOSOPHICAL ISSUES; CONSCIOUSNESS; CONSCIOUSNESS, NEUROBIOLOGY OF; DESCARTES, RENÉ; FREUD, SIGMUND; INNATENESS OF LANGUAGE; JAMES, WILLIAM; KANT, IMMANUEL; MIND-BODY PROBLEM; MODULARITY OF MIND; NATIVISM; NATIVISM, HISTORY OF; PHYSICALISM; PSYCHOANALYSIS, CONTEMPORARY VIEWS; PSYCHOANALYSIS, HISTORY OF; QUALIA; RATIONALISM VS. EMPIRICISM; SELF; SELF-KNOWLEDGE; SOCIAL PLAY BEHAVIOR; THEORY OF MIND

2   From Materialism to Mental Science

In raising issue *i.*, the mental-physical relation, in the previous section, I implied that materialism was the dominant ontological view of the mind in contemporary philosophy of mind. I also suggested that, if anything, general convergence on this issue has intensified interest in the mind-body problem. For example, consider the large and lively debate over whether contemporary forms of materialism are compatible with genuine MENTAL CAUSATION, or, alternatively, whether they commit one to EPIPHE-NOMENALISM about the mental (Kim 1993; Heil and Mele 1993; Yablo 1992). Likewise, consider the fact that despite the dominance of materialism, some philosophers maintain that there remains an EXPLANATORY GAP between mental phenomena such as consciousness and any physical story that we·are likely to get about the workings of the brain (Levine 1983; cf. Chalmers 1996). Both of these issues, very much alive in contemporary philosophy of mind and cognitive science, concern the mind-body problem, even if they are not always identified in such old-fashioned terms.

I also noted that a healthy interest in the first-person perspective persists within this general materialist framework. By taking a quick look at the two major initial attempts to develop a systematic, scientific understanding of the mind—late nineteenth-century introspectionism and early twentieth-century behaviorism—I want to elaborate on these two points and bring them together.

Introspectionism was widely held to fall prey to a problem known as the *problem of the homunculus*. Here I argue that behaviorism, too, is subject to a variation on this very problem, and that both versions of this problem continue to nag at contemporary sciences of the mind.

Students of the history of psychology are familiar with the claim that the roots of contemporary psychology can be dated from 1879, with the founding of the first experimental laboratory devoted to psychology by WILHELM WUNDT in Leipzig, Germany. As an *experimental* laboratory, Wundt's laboratory relied on the techniques introduced and refined in physiology and psychophysics over the preceding fifty years

by HELMHOLTZ, Weber, and Fechner that paid particular attention to the report of SEN-SATIONS. What distinguished Wundt's as a laboratory of *psychology* was his focus on the data reported in consciousness via the first-person perspective; psychology was to be the science of immediate experience and its most basic constituents. Yet we should remind ourselves of how restricted this conception of psychology was, particularly relative to contemporary views of the subject.

First, Wundt distinguished between mere INTROSPECTION, first-person reports of the sort that could arise in the everyday course of events, and experimentally manipulable self-observation of the sort that could only be triggered in an experimental context. Although Wundt is often thought of as the founder of an introspectionist methodology that led to a promiscuous psychological ontology, in disallowing mere introspection as an appropriate method for a science of the mind he shared at least the sort of restrictive conception of psychology with *both* his physiological predecessors and his later behaviorist critics.

Second, Wundt thought that the vast majority of ordinary thought and cognition was *not* amenable to acceptable first-person analysis, and so lay beyond the reach of a scientific psychology. Wundt thought, for example, that belief, language, personality, and SOCIAL COGNITION could be studied systematically only by detailing the cultural mores, art, and religion of whole societies (hence his four-volume *Völkerpsychologie* of 1900–1909). These studies belonged to the humanities (*Geisteswissenshaften*) rather than the experimental sciences (*Naturwissenschaften*), and were undertaken by anthropologists inspired by Wundt, such as BRONISLAW MALINOWSKI.

Wundt himself took one of his early contributions to be a solution of the mind-body problem, for that is what the data derived from the application of the experimental method to distinctly psychological phenomena gave one: correlations between the mental and the physical that indicated how the two were systematically related. The discovery of psychophysical laws of this sort showed how the mental was related to the physical. Yet with the expansion of the domain of the mental amenable to experimental investigation over the last 150 years, the mind-body problem has taken on a more acute form: just how do we get all that mind-dust from merely material mechanics? And it is here that the problem of the homunculus arises for introspectionist psychology after Wundt.

The problem, put in modern guise, is this. Suppose that one introspects, say, in order to determine the location of a certain feature (a cabin, for example) on a map that one has attempted to memorize (Kosslyn 1980). Such introspection is typically reported in terms of exploring a mental image with one's *mind's eye*. Yet we hardly want our psychological story to end there, because it posits a process (introspection) and a processor (the mind's eye) that themselves cry out for further explanation. The problem of the homunculus is the problem of leaving undischarged homunculi ("little men" or their equivalents) in one's *explanantia,* and it persists as we consider an elaboration on our initial introspective report. For example, one might well report forming a mental image of the map, and then scanning around the various features of the map, zooming in on them to discern more clearly what they are to see if any of them is the sought-after cabin. To take this introspective report seriously as a guide to the underlying psychological mechanisms would be to posit, minimally, an *imager* (to form the initial image), a *scanner* (to guide your mind's eye around the image), and a *zoomer* (to adjust the relative sizes of the features on the map). But here again we face the problem of the homunculus, because such "mechanisms" themselves require further psychological decomposition.

To be faced with the problem of the homunculus, of course, is not the same as to succumb to it. We might distinguish two understandings of just what the "problem" is here. First, the problem of the homunculus could be viewed as a problem specifically for introspectionist views of psychology, a problem that was never successfully met and that was principally responsible for the abandonment of introspectionism. As such, the problem motivated BEHAVIORISM in psychology. Second, the problem of the homunculus might simply be thought of as a challenge that *any* view that posits internal mental states must respond to: to show how to discharge all of the homunculi introduced in a way that is acceptably materialistic. So construed, the problem

remains one that has been with us more recently, in disputes over the psychological reality of various forms of GENERATIVE GRAMMAR (e.g., Stabler 1983); in the nativism that has been extremely influential in post-Piagetian accounts of COGNITIVE DEVELOPMENT (Spelke 1990; cf. Elman et al. 1996); and in debates over the significance of MENTAL ROTATION and the nature of IMAGERY (Kosslyn 1994; cf. Pylyshyn 1984: ch.8).

With Wundt's own restrictive conception of psychology and the problem of the homunculus in mind, it is with some irony that we can view the rise and fall of behaviorism as the dominant paradigm for psychology subsequent to the introspectionism that Wundt founded. For here was a view so deeply indebted to materialism and the imperative to explore psychological claims only by reference to what was acceptably experimental that, in effect, in its purest form it appeared to do away with the distinctively mental altogether! That is, because objectively observable behavioral responses to objectively measurable stimuli are all that could be rigorously explored, experimental psychological investigations would need to be significantly curtailed, relative to those of introspectionists such as Wundt and Titchener. As J. B. Watson said in his early, influential "Psychology as the Behaviorist Views It" in 1913, "Psychology as behavior will, after all, have to neglect but few of the really essential problems with which psychology as an introspective science now concerns itself. In all probability even this residue of problems may be phrased in such a way that refined methods in behavior (which certainly must come) will lead to their solution" (p. 177).

Behaviorism brought with it not simply a global conception of psychology but specific methodologies, such as CONDITIONING, and a focus on phenomena, such as that of LEARNING, that have been explored in depth since the rise of behaviorism. Rather than concentrate on these sorts of contribution to the interdisciplinary sciences of the mind that behaviorists have made, I want to focus on the central problem that faced behaviorism as a research program for reshaping psychology.

One of the common points shared by behaviorists in their philosophical and psychological guises was a commitment to an *operational* view of psychological concepts and thus a suspicion of any reliance on concepts that could not be operationally characterized. Construed as a view of scientific *definition* (as it was by philosophers), operationalism is the view that scientific terms must be defined in terms of observable and measurable operations that one can perform. Thus, an operational definition of "length," as applied to ordinary objects, might be: "the measure we obtain by laying a standard measuring rod or rods along the body of the object." Construed as a view of scientific *methodology* (as it was by psychologists), operationalism claims that the subject matter of the sciences should be objectively observable and measurable, by itself a view without much content.

The real bite of the insistence on operational definitions and methodology for psychology came via the application of operationalism to unobservables, for the various feelings, sensations, and other internal states reported by introspection, themselves unobservable, proved difficult to operationalize adequately. Notoriously, the introspective reports from various psychological laboratories produced different listings of the basic feelings and sensations that made up consciousness, and the lack of agreement here generated skepticism about the reliability of introspection as a method for revealing the structure of the mind. In psychology, this led to a focus on behavior, rather than consciousness, and to its exploration through observable stimulus and response: hence, behaviorism. But I want to suggest that this reliance on operationalism itself created a version of the problem of the homunculus for behaviorism. This point can be made in two ways, each of which offers a reinterpretation of a standard criticism of behaviorism. The first of these criticisms is usually called "philosophical behaviorism," the attempt to provide conceptual analyses of mental state terms exclusively in terms of behavior; the second is "psychological behaviorism," the research program of studying objective and observable behavior, rather than subjective and unobservable inner mental episodes.

First, as Geach (1957: chap. 4) pointed out with respect to belief, behaviorist analyses of individual folk psychological states are bound to fail, because it is only in concert with many other propositional attitudes that any given such attitude has

behavioral effects. Thus, to take a simple example, we might characterize the belief that it is raining as the tendency to utter "yes" when asked, "Do you believe that it is raining?" But one reason this would be inadequate is that one will engage in this verbal behavior only if one *wants* to answer truthfully, and only if one *hears* and *understands* the question asked, where each of the italicized terms above refers to some other mental state. Because the problem recurs in *every* putative analysis, this implies that a behavioristically acceptable construal of folk psychology is not possible. This point would seem to generalize beyond folk psychology to representational psychology more generally.

So, in explicitly attempting to do without internal mental representations, behaviorists themselves are left with mental states that must simply be assumed. Here we are not far from those undischarged homunculi that were the bane of introspectionists, especially once we recognize that the metaphorical talk of "homunculi" refers precisely to internal mental states and processes that themselves are not further explained.

Second, as Chomsky (1959: esp. p. 54) emphasized in his review of Skinner's *Verbal Behavior,* systematic attempts to operationalize psychological language invariably smuggle in a reference to the very mental processes they are trying to do without. At the most general level, the behavior of interest to the linguist, Skinner's "verbal behavior," is difficult to characterize adequately without at least an implicit reference to the sorts of psychological mechanism that generate it. For example, linguists are not interested in mere noises that have the same physical properties—"harbor" may be pronounced so that its first syllable has the same acoustic properties as an exasperated grunt—but in parts of speech that are taxonomized at least partially in terms of the surrounding mental economy of the speaker or listener.

The same seems true for *all* of the processes introduced by behaviorists—for example, stimulus control, reinforcement, conditioning—insofar as they are used to characterize complex, human behavior that has a natural psychological description (making a decision, reasoning, conducting a conversation, issuing a threat). What marks off their instances as behaviors *of the same kind* is not exclusively their physical or behavioral similarity, but, in part, the common, internal psychological processes that generate them, and that they in turn generate. Hence, the irony: behaviorists, themselves motivated by the idea of reforming psychology so as to generalize about objective, observable behavior and so avoid the problem of the homunculus, are faced with undischarged homunculi, that is, irreducibly mental processes, in their very own alternative to introspectionism.

The two versions of the problem of the homunculus are still with us as a Scylla and Charybdis for contemporary cognitive scientists to steer between. On the one hand, theorists need to avoid building the very cognitive abilities that they wish to explain into the models and theories they construct. On the other, in attempting to side-step this problem they also run the risk of masking the ways in which their "objective" taxonomic categories presuppose further internal psychological description of precisely the sort that gives rise to the problem of the homunculus in the first place.

See also BEHAVIORISM; COGNITIVE DEVELOPMENT; CONDITIONING; EPIPHENOMENALISM; EXPLANATORY GAP; GENERATIVE GRAMMAR; HELMHOLTZ, HERMANN; IMAGERY; INTROSPECTION; LEARNING; MALINOWSKI, BRONISLAW; MENTAL CAUSATION; MENTAL ROTATION; SENSATIONS; SOCIAL COGNITION; SOCIAL COGNITION IN ANIMALS; WUNDT, WILHELM

3   A Detour Before the Naturalistic Turn

Given the state of philosophy and psychology in the early 1950s, it is surprising that within twenty-five years there would be a thriving and well-focused interdisciplinary unit of study, cognitive science, to which the two are central. As we have seen, psychology was dominated by behaviorist approaches that were largely skeptical of positing internal mental states as part of a serious, scientific psychology. And Anglo-American philosophy featured two distinct trends, each of which made philosophy more insular with respect to other disciplines, and each of which served to reinforce the behaviorist orientation of psychology.

First, ordinary language philosophy, particularly in Great Britain under the influence of Ludwig Wittgenstein and J. L. Austin, demarcated distinctly philosophical problems as soluble (or dissoluble) chiefly by reference to what one would ordinarily say, and tended to see philosophical views of the past and present as the result of confusions in how philosophers and others come to use words that generally have a clear sense in their ordinary contexts. This approach to philosophical issues in the post-war period has recently been referred to by Marjorie Grene (1995: 55) as the "Bertie Wooster season in philosophy," a characterization I suspect would seem apt to many philosophers of mind interested in contemporary cognitive science (and in P. G. Wodehouse). Let me illustrate how this approach to philosophy served to isolate the philosophy of mind from the sciences of the mind with perhaps the two most influential examples pertaining to the mind in the ordinary language tradition.

In *The Concept of Mind,* Gilbert Ryle (1949: 17) attacked a view of the mind that he referred to as "Descartes' Myth" and "the dogma of the Ghost in the Machine"—basically, dualism—largely through a repeated application of the objection that dualism consisted of an extended *category mistake*: it "represents the facts of mental life as if they belonged to one logical type or category . . . when they actually belong to another." Descartes' Myth represented a category mistake because in supposing that there was a special, inner theater on which mental life is played out, it treated the "facts of mental life" as belonging to a special category of facts, when they were simply facts about how people can, do, and would behave in certain circumstances. Ryle set about showing that for the range of mental concepts that were held to refer to private, internal mental episodes or events according to Descartes' Myth—intelligence, the will, emotion, self-knowledge, sensation, and imagination—an appeal to what one would ordinarily say both shows the dogma of the Ghost in the Machine to be false, and points to a positive account of the mind that was behaviorist in orientation. To convey why Ryle's influential views here turned philosophy of mind away from science rather than towards it, consider the opening sentences of *The Concept of Mind*: "This book offers what may with reservations be described as a theory of the mind. But it does not give new information about minds. We possess already a wealth of information about minds, information which is neither derived from, nor upset by, the arguments of philosophers. The philosophical arguments which constitute this book are intended not to increase what we know about minds, but to rectify the logical geography of the knowledge which we already possess" (Ryle 1949: 9). The "we" here refers to ordinary folk, and the philosopher's task in articulating a theory of mind is to draw on what we already know about the mind, rather than on arcane, philosophical views or on specialized, scientific knowledge.

The second example is Norman Malcolm's *Dreaming,* which, like *The Concept of Mind,* framed the critique it wished to deliver as an attack on a Cartesian view of the mind. Malcolm's (1959: 4) target was the view that "dreams are the activity of the mind during sleep," and associated talk of DREAMING as involving various mental acts, such as remembering, imagining, judging, thinking, and reasoning. Malcolm argued that such dream-talk, whether it be part of commonsense reflection on dreaming (How long do dreams last?; Can you work out problems in your dreams?) or a contribution to more systematic empirical research on dreaming, was a confusion arising from the failure to attend to the proper "logic" of our ordinary talk about dreaming. Malcolm's argument proceeded by appealing to how one would *use* various expressions and sentences that contained the word "dreaming." (In looking back at Malcolm's book, it is striking that nearly every one of the eighteen short chapters begins with a paragraph about words and what one would say with or about them.)

Malcolm's central point was that there was no way to *verify* any given claim about such mental activity occurring while one was asleep, because the commonsense criteria for the application of such concepts were incompatible with saying that a person was asleep or dreaming. And because there was no way to tell whether various attributions of mental states to a sleeping person were correct, such attributions were meaningless. These claims not only could be made without an appeal to any empirical details about dreaming or SLEEP, but implied that the whole enterprise of investigating dreaming empirically itself represented some sort of *logical* muddle.

Malcolm's point became more general than one simply about dreaming (or the word "dreaming"). As he said in a preface to a later work, written after "the notion that thoughts, ideas, memories, sensations, and so on 'code into' or 'map onto' neural firing patterns in the brain" had become commonplace: "I believe that a study of our psychological concepts can show that [such] psycho-physical isomorphism is not a coherent assumption" (Malcolm 1971: x). Like Ryle's straightening of the logical geography of our knowledge of minds, Malcolm's appeal to the study of our psychological concepts could be conducted without any knowledge gleaned from psychological science (cf. Griffiths 1997: chap. 2 on the emotions).

Quite distinct from the ordinary language tradition was a second general perspective that served to make philosophical contributions to the study of the mind "distinctive" from those of science. This was logical positivism or empiricism, which developed in Europe in the 1920s and flourished in the United States through the 1930s and 1940s with the immigration to the United States of many of its leading members, including Rudolph Carnap, Hans Reichenbach, Herbert Feigl, and Carl Hempel. The logical empiricists were called "empiricists" because they held that it was via the senses and observation that we came to know about the world, deploying this empiricism with the logical techniques that had been developed by Gottlob Frege, Bertrand Russell, and Alfred Whitehead. Like empiricists in general, the logical positivists viewed the sciences as the paradigmatic repository of knowledge, and they were largely responsible for the rise of philosophy of science as a distinct subdiscipline within philosophy.

As part of their reflection on science they articulated and defended the doctrine of the UNITY OF SCIENCE, the idea that the sciences are, in some sense, essentially unified, and their empiricism led them to appeal to PARSIMONY AND SIMPLICITY as grounds for both theory choice within science and for preferring theories that were ontological Scrooges. This empiricism came with a focus on *what could be verified*, and with it scepticism about traditional metaphysical notions, such as God, CAUSATION, and essences, whose instances could not be verified by an appeal to the data of sense experience. This emphasis on verification was encapsulated in the verification theory of meaning, which held that the meaning of a sentence was its method of verification, implying that sentences without any such method were *meaningless*. In psychology, this fueled skepticism about the existence of internal mental representations and states (whose existence could not be objectively verified), and offered further philosophical backing for behaviorism.

In contrast to the ordinary language philosophers (many of whom would have been professionally embarrassed to have been caught knowing anything about science), the positivists held that philosophy was to be informed about and sensitive to the results of science. The distinctive task of the philosopher, however, was not simply to describe scientific practice, but to offer a *rational reconstruction* of it, one that made clear the logical structure of science. Although the term "*rational reconstruction*" was used first by Carnap in his 1928 book *The Logical Construction of the World,* quite a general epistemological tract, the technique to which it referred came to be applied especially to scientific concepts and theories.

This played out in the frequent appeal to the distinction between the *context of discovery* and the *context of justification,* drawn as such by Reichenbach in *Experience and Prediction* (1938) but with a longer history in the German tradition. To consider an aspect of a scientific view in the context of discovery was essentially to raise psychological, sociological, or historical questions about how that view originated, was developed, or came to be accepted or rejected. But properly philosophical explorations of science were to be conducted in the context of justification, raising questions and making claims about the logical structure of science and the concepts it used. Rational reconstruction was the chief way of divorcing the relevant scientific theory from its mere context of discovery.

A story involving Feigl and Carnap nicely illustrates the divorce between philosophy and science within positivism. In the late 1950s, Feigl visited the University of California, Los Angeles, to give a talk to the Department of Philosophy, of which Carnap was a member. Feigl's talk was aimed at showing that a form of physicalism, the