current experimental and theoretical literature in psycholinguistics. In the course of that exercise, we returned again and again to discussions of the underpinnings of the discipline. Much of what went on in those discussions is replicated here. My indebtedness to my coauthors probably verges on plagiarism, and my gratitude for what they taught me is unbounded.

Even so, this book would not have gotten written except for a sabbatical grant from M.I.T. and a concurrent fellowship from the Guggenheim Foundation, which, together, freed me from academic duties during the year 1973–1974. My obligation to both institutions is hereby gratefully acknowledged.

I tried out early versions of some of the material in this book in series of lectures at the Department of Psychology at the University of Oxford and the Department of Philosophy at University College, University of London. I should like to offer my thanks to Dr. Ann Treisman for arranging the former lectures and to Professor Richard Wollheim for arranging the others; also to the students and faculty at both institutions for providing useful comments and criticisms.

Finally, a number of friends and relations have read all or parts of the manuscript, invariably to good effect. Alas, none of the following are responsible for the residual errors: Professors Ned Block, Susan Carey Block, George Boolos, Noam Chomsky, Janet Dean Fodor, Jerrold Katz, Edward Martin, and George Miller. I am especially obliged to Mr. Georges Rey, who read the manuscript with great care and provided invaluable criticism and advice; and to Mrs. Cornelia Parkes, who helped with the bibliography.

The second half of the Introduction to this book is a version, slightly revised, of a paper called "Special Sciences," which first appeared in *Synthese* (Fodor, 1974). Permission to republish this material is gratefully acknowledged. Material quoted from Chapters one and two and the conclusion of *The Construction of Reality in the Child* by Jean Piaget (translated by Marjorie Cooke) is copyrighted 1954 by Basic Books, Inc. Publishers, New York, and is used with their permission. Other quoted materials are used with the permission of D. Riedel Publishing Co.; Penguin Books Ltd.; John Wiley and Sons Inc.; The Humanities Press Inc.; and Routledge and Kegen Paul, Ltd.

# INTRODUCTION: TWO KINDS OF REDUCTIONISM

AICS4 STUDENTS

IPSO FACTO: BY THAT VERY FACT

"mind different from brain"

*The man who laughs is the one who has not yet heard the terrible news.*
BERTHOLD BRECHT

I propose, in this book, to discuss some aspects of the theory of mental processes. Many readers may, however, feel that this choice of topic is ill-advised: either because they think there are no such processes to discuss or because they think there is no theory about them whose aspects will bear discussing. The second of these worries is substantive, and its consideration must be deferred to the body of the text. The best demonstration that speculative psychology can be done is, after all, to do some. But I am aware that the distrust with which many philosophers, and many philosophically sophisticated psychologists, view the kind of inquiry I shall undertake stems from something more than a jaundiced appreciation of the empirical literature. It is with the sources of this suspicion that the present chapter will primarily be concerned.

The integrity of psychological theorizing has always been jeopardized by two kinds of reductionism, each of which would vitiate the psychologist's claim to study mental phenomena. For those influenced by the tradition of logical behaviorism, such phenomena are allowed no ontological status distinct from the behavioral events that psychological theories explain. Psychology is thus deprived of its theoretical terms except where these can be construed as nonce locutions for which behavioral reductions will eventually be provided. To all intents and purposes, this means that psychologists can provide methodologically reputable accounts only of such aspects of behavior as are the effects of environmental variables.

make faulty

science of pure being

for the time being

Not surprisingly, many psychologists have found this sort of methodology intolerably restrictive: The contribution of the organism's internal states to the causation of its own behavior seems sufficiently undisputable, given the spontaneity and freedom from local environmental control that

2

behavior often exhibits. Behaviorism thus invites us to deny the undisputable, but, in fact, we need not do so; there is an alternative that frequently gets endorsed. We can acknowledge that behavior is largely the effect of organic processes so long as we bear in mind that these processes *are* organic: i.e., that they are physiological processes located, presumably, in the nervous systems of organisms. Psychology can thus avoid behavioral reduction by opting for physiological reduction, but it must opt one way or the other.

Either way, the psychologist loses. Insofar as psychological explanations are allowed a theoretical vocabulary, it is the vocabulary of some *different* science (neurology or physiology). Insofar as there *are* laws about the ways in which behavior is contingent upon internal processes, it is the neurologist or the physiologist who will, in the long run, get to state them. However psychologists choose between the available reductions, their discipline is left without a proprietary subject matter. The best a working psychologist can hope for is an interim existence eked out between the horns of this dilemma and (just) tolerated by colleagues in the 'hard' sciences.

their own

I think, however, that this is a false dilemma. I know of no convincing reason why a science should not seek to exhibit the contingency of an organism's behavior upon its internal states, and I know of no convincing reason why a science which succeeds in doing so should be reducible to brain science; not, at least, in the sense of reduction which would entail that psychological theories can somehow be *replaced* by their physiological counterparts. I shall try, in this introductory chapter, to show that both horns of the dilemma are, in fact, blunt. By doing so I hope to undermine a number of the arguments that are usually alleged against types of psychological explanations which, in succeeding chapters, I shall be taking very seriously indeed.

Not a science as can't study beliefs etc as internal. Pavlov's dog (Bell) study with respect to external behaviour eg learning. Internal states irrelevant.

## LOGICAL BEHAVIORISM

Many philosophers, and some scientists, seem to hold that the sorts of theories now widely endorsed by cognitive psychologists could not conceivably illuminate the character of mental processes. For, it is claimed, such theories assume a view of psychological explanation which is, and has been shown to be, fundamentally incoherent. The line, to put it crudely, is that Ryle and Wittgenstein killed this sort of psychology some time about 1945, and there is no point to speculating on the prospects of the deceased.

I shall not attempt a full-dress refutation of this view. If the Wittgensteinian tradition in the philosophy of mind does, indeed, offer a coherent attack upon the methodology of current cognitive psychology, it is one which depends on a complex of assumptions about the nature of explanation, the ontological status of theoretical entities, and the a priori conditions upon the possibility of linguistic communication. To meet that attack head on would require showing—what, in fact, I believe is true—that these assumptions, insofar as they are clear, are unwarranted. But that is a book's work in itself, and not a book that I feel much like writing. The best that I can do here is to sketch a preliminary defense of the methodological commitments implicit in the kind of psychological theorizing with which I shall be mainly concerned. Insofar as these commitments differ from what many philosophers have been willing to accept, even a sketch of their defense may prove to be revealing.

Among the many passages in Ryle's *Concept of Mind* (1949) that repay close attention, there is one (around p. 33) in which the cards are more than usually on the table. Ryle is discussing the question: 'What makes a clown's clowning intelligent (witty, clever, ingenious, etc.)?' The doctrine he is disapproving goes as follows: What makes the clowning intelligent is the fact that it is the consequence of certain mental operations (computations, calculations) privy to the clown and causally responsible for the production of the clown's behavior. Had these operations been other than they were, then (the doctrine claims) either the clowning would have been witless or at least it would have been witty clowning of some different kind. In short, the clown's clowning was clever in the way that it was because the mental operations upon which the clowning was causally contingent had whatever character they did have. And, though Ryle doesn't say so, it is presumably implied by this doctrine that a psychologist interested in explaining the success of the clown's performance would ipso facto be in the business of saying what those operations were and how, precisely, they were related to the overt pratfalls that the crowd saw.

Strictly speaking, this is not a single theory but a batch of closely connected ones. In particular, one can distinguish at least three claims about the character of the events upon which the clown's behavior is said to be causally contingent:

1. That some of them are mental events;
2. That some (or all) of the mental events are privy to the clown in at least the sense that they are normally unobserved by someone who observes the clown's performance, and, perhaps, also in the stronger sense that they are in principle unobservable by anyone except the clown;
3. That it is the fact that the behavior was caused by such events that makes it the kind of behavior it is; that intelligent behavior *is* intelligent because it has the kind of etiology it has. Study of Causation

I want to distinguish these doctrines because a psychologist might accept the sorts of theories that Ryle doesn't like without wanting to commit himself to the full implications of what Ryle calls 'Cartesianism'. For exam-

Good/evil. mind/body Decarte · laws for each.

3

ple, Ryle assumes (as most psychologists who take a Realistic view of the designata of mental terms in psychological theories would not) that a mentalist must be a dualist; in particular, that mentalism and materialism are mutually exclusive. I have argued elsewhere that confusing mentalism with dualism is the original sin of the Wittgensteinian tradition (cf. Fodor, 1968, especially Chap. 2). Suffice it to remark here that one result of this confusion is the tendency to see the options of dualism and behaviorism as exhaustive in the philosophy of mind.

spirit / matter

Similarly, it seems to me, one might accept some such view as that of item 3 without embracing a doctrinaire reading of item 2. It may be that some of the mental processes that are causally responsible for the clown's behavior are de facto unobservable by the crowd. It may be, for that matter, that some of these processes are de facto unobservable by the clown. But there would seem to be nothing in the project of explaining behavior by reference to mental processes which requires a commitment to epistemological privacy in the traditional sense of that notion. Indeed, for better or for worse, a materialist *cannot* accept such a commitment since his view is that mental events are species of physical events, and physical events are publicly observable at least in principle.[1, 2]

It is notorious that, even granting these caveats, Ryle doesn't think this kind of account could possibly be true. For this theory says that what makes the clown's clowning clever is the fact that it is the effect of a certain kind of cause. But what, in Ryle's view, actually *does* make the clowning clever is something quite else: For example, the fact that it happens out where the audience can see it; the fact that the things that the clown does are not the things that the audience expected him to do; the fact that the man he hit with the pie was dressed in evening clothes, etc.

There are two points to notice. First, none of *these* facts are in any sense private to the clown. They are not even de facto private in the sense of being facts about things going on in the clown's nervous system. On the contrary, what makes the clown's clowning clever is precisely the *public* aspects of his performance; precisely the things that the audience *can* see. The second point is that what makes the clowning clever is not the character of the *causes* of the clown's behavior, but rather the character of the behavior itself. It counts for the pratfall being clever that it occurred when it wasn't expected, but its occurring when it wasn't expected surely wasn't one of its causes on any conceivable construal of 'cause'. In short, what makes the clowning clever is not some event distinct from, and causally responsible for, the behavior that the clown produces. A fortiori, it is not a mental event prior to the pratfall. Surely, then, if the mentalist program involves the identification and characterization of such an event, that program is doomed from the start.

Alas for the psychology of clever clowning. We had assumed that psychologists would identify the (mental) causes upon which clever clowning is contingent and *thereby* answer the question: 'What makes the clowning clever?' Now all that appears to be left of the enterprise is the alliterations. Nor does Ryle restrict his use of this pattern of argument to undermining the psychology of clowns. Precisely similar moves are made to show that the psychology of perception is a muddle since what makes something (e.g.) the recognition of a robin or a tune is not the occurrence of some or other mental event, but rather the fact that what was claimed to be a robin was in fact a robin, and what was taken to be a rendition of "Lillibullero" was one. It is, in fact, hard to think of an area of cognitive psychology in which this sort of argument would not apply or where Ryle does not apply it. Indeed, it is perhaps Ryle's *central* point that 'Cartesian' (i.e., mentalistic) psychological theories treat what is really a *logical* relation between aspects of a single event as though it were a causal relation between pairs of distinct events. It is this tendency to give mechanistic answers to conceptual questions which, according to Ryle, leads the mentalist to orgies of regrettable hypostasis: i.e., to attempting to explain behavior by reference to underlying psychological mechanisms.[3]

dependent on

essential substance.

[1] The purist will note that this last point depends on the (reasonable) assumption that the context 'is publicly observable at least in principle' is transparent to substitutivity of identicals.

[2] It might be replied that if we allow the possibility that mental events might be physical events, that some mental events might be unconscious, and that no mental event is essentially private, we will have so attenuated the term 'mental' as to deprive it of all force. It is, of course, true that the very notion of a mental event is often specified in ways that presuppose dualism and/or a strong doctrine of epistemological privacy. What is unclear, however, is what we want a definition of 'mental event' *for* in the first place.

Surely not, in any event, in order that it should be possible to do psychology in a methodologically respectable way. *Pre*-theoretically we identify mental events by reference to clear cases. *Post*-theoretically it is sufficient to identify them as the ones which fall under psychological laws. This characterization is, of course, question-begging since it rests upon an unexplained distinction between psychological laws and all the others. The present point, however, is that we are in no better position vis-à-vis such notions as chemical event (or meteorological event, or geological event . . . , etc.), a state of affairs which does not prejudice the rational pursuit of chemistry. A chemical event is one that falls under chemical laws; chemical laws are those which follow from (ideally completed) chemical theories; chemical theories are theories in chemistry; and chemistry, like all other special sciences, is individualated large post facto and by reference to its typical problems and predicates. (For example, chemistry is that science which concerns itself with such matters as the combinatorial properties of elements, the analysis and synthesis of compounds, etc.) Why, precisely, is this not good enough?

[3] 'Criterion' isn't one of Ryle's words: Nevertheless, the line of argument just reviewed relates Ryle's work closely to the criteriological tradition in post-Wittgensteinian philosophy of mind. Roughly, what in Ryle's terms "makes" *a* be *F* is *a's* possession of those properties which are criterial for the application of '*F*' to *xs*.

to pass through.

If this *is* a mistake I am in trouble. For it will be the pervasive assumption of my discussion that such explanations, however often they may prove to be empirically unsound, are, in principle, methodologically impeccable. What I propose to do throughout this book is to take such explanations absolutely seriously and try to sketch at least the outlines of the general picture of mental life to which they lead. So something will have to be done to meet Ryle's argument. Let's, to begin with, vary the example.

Consider the question: 'What makes Wheaties the breakfast of champions?' (Wheaties, in case anyone hasn't heard, is, or are, a sort of packaged cereal. The details are very inessential.) There are, it will be noticed, at least two kinds of answers that one might give.[4] A sketch of one answer, which belongs to what I shall call the 'causal story' might be: 'What make Wheaties the breakfast of champions are the health-giving vitamins and minerals that it contains'; or 'It's the carbohydrates in Wheaties, which give one the energy one needs for hard days on the high hurdle'; or 'It's the special springiness of all the little molecules in Wheaties, which gives Wheaties eaters their unusually high coefficient or restitution', etc.

It's not important to my point that any of these specimen answers should be true. What *is* essential is that some causal story or other must be true if Wheaties really *are* the breakfast of champions as they are claimed to be. Answers propose causal stories insofar as they seek to specify properties of Wheaties which may be causally implicated in the processes that make champions of Wheaties eaters. Very roughly, such answers suggest provisional values of $P$ in the explanation schema: '$P$ causes (($x$ eats Wheaties) brings about ($x$ becomes a champion)) for significantly many values of $x$'. I assume that, if Wheaties *do* make champions of those who eat them, then there must be at least one value of $P$ which makes this schema true. Since that assumption is simply the denial of the miracle theory of Wheaties, it ought not be in dispute.

[4] I am reading 'What makes Wheaties the breakfast of champions?' as asking 'What about Wheaties makes champions of (some, many, so many) Wheaties eaters?' rather than 'What about Wheaties makes (some, many, so many) champions eat them?' The latter question invites the reasons that champions give for eating Wheaties; and though these *may* include reference to properties Wheaties have by virtue of which its eaters become champions, they need not do so. Thus, a plausible answer to the second question which is *not* plausibly an answer to the first might be: 'They taste good'.

I am uncertain which of these questions the Wheaties people have in mind when they ask 'What makes Wheaties the breakfast of champions?' rhetorically, as, I believe, they are wont to do. Much of their advertising consists of publicizing statements by champions to the effect that they (the champions) do, in fact, eat Wheaties. If, as may be the case, such statements are offered as arguments for the truth of the presupposition of the question on its *first* reading (viz., that there is something about Wheaties that makes champions of those who eat them), then it would appear that General Mills has either misused the method of differences or committed the fallacy



4

I suggested that there is another kind of answer that 'What makes Wheaties the breakfast of champions?' may appropriately receive. I will say that answers of this second kind belong to the 'conceptual story'. In the present case, we can tell the conceptual story with some precision: What makes Wheaties the breakfast of champions is the fact that it is eaten (for breakfast) by nonnegligible numbers of champions. This is, I take it, a conceptually necessary and sufficient condition for *anything* to be the breakfast of champions;[5] as such, it pretty much exhausts the conceptual story about Wheaties.

The point to notice is that answers that belong to the conceptual story typically do not belong to the causal story and vice versa.[6] In particular, its being eaten by nonnegligible numbers of champions does not *cause* Wheaties to be the breakfast of champions; no more than its occurring unexpectedly causes the clown's pratfall to be witty. Rather, what we have in both cases are instances of (more or less rigorous) conceptual connections. Being eaten by nonnegligible numbers of champions and being unexpected belong, respectively, to the analyses of 'being the breakfast of champions' and 'being witty', with the exception that, in the former case, we have something that approaches a logically necessary and sufficient condition and, in the latter case, we very clearly do not.[7]

The notion of conceptual connection is notoriously a philosophical miasma; all the more so if one holds (as Wittgensteinians usually do) that there are species of conceptual connections which cannot, even in principle,

an unwholesome exhalation.

[5] This is not quite right. Being eaten for breakfast by nonnegligible numbers of champions is a conceptually necessary and sufficient condition for something being *a* breakfast of champions (cf. Russell, 1905). Henceforth I shall resist this sort of pedantry whenever I can bring myself to do so.

[6] The exceptions are interesting. They involve cases where the conceptual conditions for something being a thing of a certain kind include the requirement that it have, or be, a certain kind of cause. I suppose, for example, that it is a conceptual truth that nothing counts as a drunken brawl unless the drunkness of the brawlers contributed causally to bringing about the brawling. See also: flu viruses, tears of rage, suicides, nervous stammers, etc. Indeed, one can imagine an analysis of 'the breakfast of champions' which would make it one of these cases too; viz. an analysis which says that it is logically necessary that the breakfast of champions is (not only what champions eat for breakfast but also) what champions eat for breakfast that is causally responsible for their being champions. But enough!

[7] It is, by the way, no accident that the latter analysis is incomplete. The usual situation is that the logically necessary and sufficient conditions for the ascription of a mental state to an organism refer not just to environmental variables but to other mental states of that organism. (For example, to *know* that $P$ is to *believe* that $P$ and to satisfy certain further conditions; to be *greedy* is to be disposed *to feel pleasure* at getting, or at the prospect of getting, more than one's share, etc.) The faith that there *must* be a way out of this network of interdependent mental terms—that one will surely get to pure behavioral ascriptions if only one pursues the analysis far enough—is, so far as I know, unsupported by either argument or example.

5

be explicated in terms of the notions of logically necessary and/or sufficient conditions. The present point, however, is that on *any* reasonable construal of conceptual connectedness, Wheaties prove that *both* the causal *and* the conceptual story can be simultaneously true, distinct answers to questions of the form: 'What makes (an) $x$ (an) $F$?' To put it succinctly, the dietitian who appears on television to explain that Wheaties is the breakfast of champions because it contains vitamins is not refuted by the philosopher who observes (though not, usually, on television) that Wheaties is the breakfast of champions because champions eat it for breakfast. The dietitian, in saying what he says, does not suppose that his remarks express, or can replace, the relevant conceptual truths. The philosopher, in saying what *he* says, ought not suppose that his remarks express, or can replace, the relevant causal explanations.

In general, suppose that $C$ is a conceptually sufficient condition for having the property $P$, and suppose that some individual $a$ does, in brute fact, satisfy $C$, so that '$Pa$' is a contingent statement true of $a$. Then: (a) it is normally pertinent to ask for a causal/mechanistic explanation of the fact that '$Pa$' *is* true; (b) such an explanation will normally constitute a (candidate) answer to the question: 'What makes $a$ exhibit the property $P$?'; (c) referring to the fact that $a$ satisfies $C$ will normally *not* constitute a causal/mechanistic explanation of the fact that $a$ exhibits the property $P$; although, (d) references to the fact that $a$ satisfies $C$ may constitute a certain (different) kind of answer to 'What makes '$Pa$' true?' I take it that, barring the looseness of the notion of a conceptual connection (and, for that matter, the looseness of the notion of a causal explanation) this pattern applies in the special case where $C$ is the property of being unexpected, $a$ is a pratfall, and '$Pa$' is the statement that $a$ was witty.

To put this point as generally as I know how, even if the behaviorists were right in supposing that logically necessary and sufficient conditions for behavior being of a certain kind can be given (just) in terms of stimulus and response variables, that fact would not in the least prejudice the mentalist's claim that the *causation* of behavior is determined by, and explicable in terms of, the organism's internal states. So far as I know, the philosophical school of 'logical' behaviorism offers not a shadow of an argument for believing that this claim is false. And the failure of behavioristic psychology to provide even a first approximation to a plausible theory of cognition suggests that the mentalist's claim may very well be true.

The arguments we have been considering are directed against a kind of reductionism which seeks to show, somehow or other, that the mental events that psychological explanations appeal to cannot be causal antecedents of the behavioral events that psychological theories seek to account for; a fortiori that statements which attribute the intelligence of a performance to the quality of the agent's cerebrations can't be [illegible]ological. The recurrent theme in this sort of reductionism is the allegation of a conceptual

action of the brain

connection between the behavioral and the mental predicates in typical instances of psychological explanations. It is from the existence of this connection that the second-class ontological status of mental events is inferred.

It should be clear by now that I don't think that this sort of argument will go through. I shall therefore assume, in what follows, that psychologists are typically in the business of supplying theories about the events that causally mediate the production of behavior and that cognitive psychologists are typically in the business of supplying theories about the events that causally mediate the production of intelligent behavior. There is, of course, no guarantee that this game can be played. It is quite conceivable that the kinds of concepts in terms of which current psychological theories are elaborated *will* turn out, in the long run, to be unsuitable for the explanation of behavior. It is, for that matter, quite conceivable that the mental processes which mediate the production of behavior are just too complicated for anyone to understand. One never can show, a priori, that a program of empirical research will certainly prove fruitful. My point has been only that the logical behaviorists have provided no a priori reason to suppose that the mentalist program in psychology will not.

Still, if mental events aren't to be reduced to behavioral events, what *are* we to say about their ontological status? I think it is very likely that all of the organismic causes of behavior are physiological, hence that mental events have true descriptions in the vocabulary of an ideally completed physiology. But I do not think that it is interesting that I think this. In particular, I don't suppose that it even begins to follow from this sort of materialism that any branch of physiology does or could supply the appropriate vocabulary for the construction of psychological theories. The likelihood that psychological events are physiological events does not entail the reducibility of psychology to physiology, ever so many philosophers and physiologists to the contrary notwithstanding. To see why this is so requires a fairly extensive discussion of the whole idea of interscience reduction, a notion which has done as much to obscure the methodology of psychology as any other except, perhaps, the verifiability criterion of meaning.

What the heck is this all about?

## PHYSIOLOGICAL REDUCTIONISM

A typical thesis of positivistic philosophy of science is that all true theories in the special sciences should reduce to physical theories 'in the long run'. This is intended to be an empirical thesis, and part of the evidence which supports it is provided by such scientific successes as the molecular theory of heat and the physical explanation of the chemical bond. But the philosophical popularity of the reductionist program cannot be explained by reference to the achievements alone. The development of science has witnessed the proliferation of specialized disciplines at least as often as it has witnessed their elimination, so the widespread enthusiasm for the view that there will eventually be only physics can hardly be a mere induction over past reductionist successes.

I think that many philosophers who accept reductionism do so primarily because they wish to endorse the generality of physics vis-à-vis the special sciences: roughly, the view that all events which fall under the laws of any science are physical events and hence fall under the laws of physics.[8] For such philosophers, saying that physics is basic science and saying that theories in the special sciences must reduce to physical theories have seemed to be two ways of saying the same thing, so that the latter doctrine has come to be a standard construal of the former.

In what follows, I shall argue that this is a considerable confusion. What has traditionally been called 'the unity of science' is a much stronger, and much less plausible, thesis than the generality of physics. If this is true it is important. Though reductionism is an empirical doctrine, it is intended to play a regulative role in scientific practice. Reducibility to physics is taken to be a *constraint* upon the acceptability of theories in the special sciences, with the curious consequence that the more the special sciences succeed, the more they ought to disappear. Methodological problems about psychology, in particular, arise in just this way: The assumption that the subject matter of psychology is part of the subject matter of physics is taken to imply that psychological theories must reduce to physical theories, and it is this latter principle that makes the trouble. I want to avoid the trouble by challenging the inference.

Reductionism is the view that all the special sciences reduce to physics. The sense of 'reduce to' is, however, proprietary. It can be characterized as follows.[9]

Let formula (1) be a law of the special science $S$.

(1) $S_1x \rightarrow S_2y$

Formula (1) is intended to be read as something like 'all events which consist of $x$'s being $S_1$ bring about events which consist of $y$'s being $S_2$'. I

[8] For expository convenience, I shall usually assume that sciences are about events in at least the sense that it is the occurrence of events that makes the laws of a science true. Nothing, however, hangs on this assumption.

[9] The version of reductionism I shall be concerned with is a stronger one than many philosophers of science hold, a point worth emphasizing since my argument will be precisely that it is too strong to get away with. Still, I think that what I shall be attacking is what many people have in mind when they refer to the unity of science, and I suspect (though I shan't try to prove it) that many of the liberalized versions of reductionism suffer from the same basic defect as what I shall take to be the classical form of the doctrine.

individualized



assume that a science is individuated largely by reference to its typical predicates (see footnote 2 above), hence that if $S$ is a special science '$S_1$' and '$S_2$' are not predicates of basic physics. (I also assume that the 'all' which quantifies laws of the special sciences needs to be taken with a grain of salt. Such laws are typically *not* exceptionless. This is a point to which I shall return at length.) A necessary and sufficient condition for the reduction of formula (1) to a law of physics is that the formulae (2) and (3) should be laws, and a necessary and sufficient condition for the reduction

(2a) $S_1x \rightleftarrows P_1x$
(2b) $S_2y \rightleftarrows P_2y$
(3) $P_1x \rightarrow P_2y$

of $S$ to physics is that all its laws should be so reduced.[10]

'$P_1$' and '$P_2$' are supposed to be predicates of physics, and formula (3) is supposed to be a physical law. Formulae like (2) are often called 'bridge' laws. Their characteristic feature is that they contain predicates of both the reduced and the reducing science. Bridge laws like formula (2) are thus contrasted with 'proper' laws like formulae (1) and (3). The upshot of the remarks so far is that the reduction of a science requires that any formula which appears as the antecedent or consequent of one of its proper laws must appear as the reduced formula in some bridge law or other.[11]

Several points about the connective '$\rightarrow$' are now in order. First, whatever properties that connective may have, it is universally agreed that it must be transitive. This is important because it is usually assumed that the reduction of some of the special sciences proceeds via bridge laws which connect their predicates with those of intermediate reducing theories. Thus, psychology is presumed to reduce to physics via, say, neurology, biochemistry, and other local stops. The present point is that this makes no difference to the logic of the situation so long as the transitivity of '$\rightarrow$' is assumed. Bridge laws which connect the predicates of $S$ to those of $S^*$ will satisfy the constraints upon the reduction of $S$ to physics so long as there are other bridge laws which, directly or indirectly, connect the predicates of $S^*$ to physical predicates.

There are, however, quite serious open questions about the interpreta-

[10] There is an implicit assumption that a science simply *is* a formulation of a set of laws. I think that this assumption is implausible, but it is usually made when the unity of science is discussed, and it is neutral so far as the main argument of this chapter is concerned.

[11] I shall sometimes refer to 'the predicate which constitutes the antecedent or consequent of a law'. This is shorthand for 'the predicate such that the antecedent or consequent of a law consists of that predicate, together with its bound variables and the quantifiers which bind them'. (Truth functions of elementary predicates are, of course, themselves predicates in this usage.)

7

tion of '→' in bridge laws. What turns on these questions is the extent to which reductionism is taken to be a physicalist thesis.

To begin with, if we read '→' as 'brings about' or 'causes' in proper laws, we will have to have some other connective for bridge laws, since bringing about and causing are presumably *asymmetric*, while bridge laws express symmetric relations. Moreover, unless bridge laws hold by virtue of the *identity* of the events which satisfy their antecedents with those that satisfy their consequents, reductionism will guarantee only a weak version of physicalism, and this would fail to express the underlying ontological bias of the reductionist program.

If bridge laws are not identity statements, then formulae like (2) claim at most that, by law, $x$'s satisfaction of a $P$ predicate and $x$'s satisfaction of an $S$ predicate are causally correlated. It follows from this that it is nomologically necessary that $S$ and $P$ predicates apply to the same things (i.e., that $S$ predicates apply to a subset of the things that $P$ predicates apply to). But, of course, this is compatible with a nonphysicalist ontology since it is compatible with the possibilty that $x$'s satsfying $S$ should not itself be a physical event. On this interpretation, the truth of reductionism does *not* guarantee the generality of physics vis-à-vis the special sciences since there are some events (satisfactions of $S$ predicates) which fall in the domains of a special science ($S$) but not in the domain of physics. (One could imagine, for example, a doctrine according to which physical and psychological predicates are both held to apply to organisms, but where it is denied that the event which consists of an organism's satisfying a psychological predicate is, in any sense, a physical event. The upshot would be a kind of psychophysical dualism of a non-Cartesian variety; a dualism of events and/or properties rather than substances.)

Given these sorts of considerations, many philosophers have held that bridge laws like formula (2) ought to be taken to express contingent event identities, so that one would read formula (2a) in some such fashion as 'every event which consists of an $x$'s satisfying $S_1$ is identical to some event which consists of that $x$'s satisfying $P_1$ and vice versa'. On this reading, the truth of reductionism would entail that every event that falls under any scientific law is a physical event, thereby simultaneously expressing the ontological bias of reductionism and guaranteeing the generality of physics vis-à-vis the special sciences.

If the bridge laws express event identities, and if every event that falls under the proper laws of a special science falls under a bridge law, we get classical reductionism, a doctrine that entails the truth of what I shall call 'token physicalism'. Token physicalism is simply the claim that all the events that the sciences talk about are physical events. There are three things to notice about token physicalism. XXX

① First, it is weaker than what is usually called 'materialism'. Materialism claims *both* that token physicalism is true *and* that every event falls under the laws of some science or other. One could therefore be a token physicalist without being a materialist, though I don't see why anyone would bother.

② Second, token physicalism is weaker than what might be called 'type physicalism', the doctrine, roughly, that every *property* mentioned in the laws of any science is a physical property. Token physicalism does not entail type physicalism, if only because the contingent identity of a pair of events presumably does not guarantee the identity of the properties whose instantiation constitutes the events; not even when the event identity is nomologically necessary. On the other hand, if an event is simply the instantiation of a property, then type physicalism does entail token physicalism; two events will be identical when they consist of the instantiation of the same property by the same individual at the same time.

③ Third, token physicalism is weaker than reductionism. Since this point is, in a certain sense, the burden of the argument to follow, I shan't labor it here. But, as a first approximation, reductionism is the conjunction of token physicalism with the assumption that there are natural kind predicates in an ideally completed physics which correspond to each natural kind predicate in any ideally completed special science. It will be one of my morals that reductionism cannot be inferred from the assumption that token physicalism is true. Reductionism is a sufficient, but not a necessary, condition for token physicalism. XXX

To summarize: I shall be reading reductionism as entailing token physicalism since, if bridge laws state nomologically necessary contingent event identities, a reduction of psychology to neurology would require that any event which consists of the instantiation of a psychological property is identical with some event which consists of the instantiation of a neurological property. Both reductionism and token physicalism entail the generality of physics, since both hold that any event which falls within the universe of discourse of a special science will also fall within the universe of discourse of physics. Moreover, it is a consequence of both doctrines that any prediction which follows from the laws of a special science (and a statement of initial conditions) will follow equally from a theory which consists only of physics and the bridge laws (together with the statement of initial conditions). Finally, it is assumed by both reductionism and token physicalism that physics is the *only* basic science; *viz.* that it is the only science that is general in the senses just specified.

Psychology
↓
Physiology
Chemistry & Biology
↓
Physics

I now want to argue that reductionism is too strong a constraint upon the unity of science, but that, for any reasonable purposes, the weaker doctrine will do.

I

Every science implies a taxonomy of the events in its universe of discourse. In particular, every science employs a descriptive vocabulary of theoretical and observation predicates, such that events fall under the laws of the science by virtue of satisfying those predicates. Patently, not every classification true description of an event is a description in such a vocabulary. For example, there are a large number of events which consist of things having been transported to a distance of less than three miles from the Eiffel Tower. I take it, however, that there is no science which contains 'is transported to a distance of less than three miles from the Eiffel Tower' as part of its descriptive vocabulary. Equivalently, I take it that there is no natural law which applies to events in virtue of their instantiating the property *is transported to a distance of less than three miles from the Eiffel Tower* (though I suppose it is just conceivable that there is some law that applies to events in virtue of their instantiating some distinct but coextensive property). By way of abbreviating these facts, I shall say that the property *is transported* . . . does not determine a (*natural*) *kind*, and that predicates which express that property are not (natural) kind predicates.

If I knew what a law is, and if I believed that scientific theories consist just of bodies of laws, then I could say that '*P*' is a kind predicate relative to *S* iff *S* contains proper laws of the form '$P_x \rightarrow \ldots y$' or '$\ldots y \rightarrow P_x$': roughly, the kind predicates of a science are the ones whose terms are the bound variables in its proper laws. I am inclined to say this even in my present state of ignorance, accepting the consequence that it makes the murky notion of a kind viciously dependent on the equally murky notions of *law* and *theory*. There is no firm footing here. If we disagree about what a kind is, we will probably also disagree about what a law is, and for the same reasons. I don't know how to break out of this circle, but I think that there are some interesting things to say about which circle we are in.

For example, we can now characterize the respect in which reductionism is too strong a construal of the doctrine of the unity of science. If reductionism is true, then *every* kind is, or is coextensive with, a physical kind. (Every kind *is* a physical kind if bridge statements express nomologically necessary property identities, and every kind is coextensive with a physical kind if bridge statements express nomologically necessary event identities.) This follows immediately from the reductionist premise that every predicate which appears as the antecedent or consequent of a law of a special science must appear as one of the reduced predicates in some bridge law, together with the assumption that the kind predicates are the ones whose terms are the bound variables in proper laws. If, in short, some physical law is related to each law of a special science in the way that formula (3) is related to formula (1), then every kind predicate of a special science is related to a kind predicate of physics in the way that formula (2) relates '$S_1$' and '$S_2$' to '$P_1$' and '$P_2$' respectively.

I now want to suggest some reasons for believing that this consequence is intolerable. These are not supposed to be knock-down reasons; they couldn't be, given that the question of whether reductionism is too strong is finally an *empirical* question. (The world could turn out to be such that every kind corresponds to a physical kind, just as it could turn out to be such that the property *is transported to a distance of less than three miles from the Eiffel Tower* determines a kind in, say, hydrodynamics. It's just that, as things stand, it seems very unlikely that the world *will* turn out to be either of these ways.)

The reason it is unlikely that every kind corresponds to a physical kind is just that (a) interesting generalizations (e.g., counterfactual supporting generalizations) can often be made about events whose physical descriptions have nothing in common; (b) it is often the case that *whether* the physical descriptions of the events subsumed by such generalizations have anything in common is, in an obvious sense, entirely irrelevant to the truth of the generalizations, or to their interestingness, or to their degree of confirmation, or, indeed, to any of their epistemologically important properties; and (c) the special sciences are very much in the business of formulating generalizations of this kind.

I take it that these remarks are obvious to the point of self-certification; they leap to the eye as soon as one makes the (apparently radical) move of taking the existence of the special sciences at all seriously. Suppose, for example, that Gresham's 'law' really is true. (If one doesn't like Gresham's law, then any true and counterfactual supporting generalization of any conceivable future economics will probably do as well.) Gresham's law says something about what will happen in monetary exchanges under certain conditions. I am willing to believe that physics is general *in the sense that it implies that any event which consists of a monetary exchange* (hence any event which falls under Gresham's law) *has a true description in the vocabulary of physics and in virtue of which it falls under the laws of physics*. But banal considerations suggest that a physical description which covers all such events must be wildly disjunctive. Some monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name to a check. What are the chances that a disjunction of physical predicates which covers all these events (i.e., a disjunctive predicate which can form the right hand side of a bridge law of the form '*x* is a monetary exchange $\rightleftharpoons \ldots$') expresses a physical kind? In particular, what are the chances that such a predicate forms the antecedent or consequent of some proper law of physics? The point is that monetary exchanges have interesting things in common; Gresham's law, if true, says what one of these interesting things is. But what is interesting about monetary exchanges is surely not their commonalities under *physical* description. A kind like a monetary exchange *could* turn out to be coextensive with a physical kind; but if it did, that would be an accident on a cosmic scale.

In fact, the situation for reductionism is still worse than the discussion thus far suggests. For reductionism claims not only that all kinds are coextensive with physical kinds, but that the coextensions are nomologically necessary: bridge laws are *laws*. So, if Gresham's law is true, it follows that there is a (bridge) law of nature such that '*x* is a monetary exchange $\rightleftharpoons$ *x*'

8

is $P'$ is true for every value of $x$, and such that $P$ is a term for a physical kind. But, surely, there is no such law. If there were, then $P$ would have to cover not only all the systems of monetary exchange that there *are*, but also all the systems of monetary exchange that there *could be*; a law must succeed with the counterfactuals. What physical predicate is a candidate for $P$ in '$x$ is a nomologically possible monetary exchange iff $P_x$'?

To summarize: An immortal econophysicist might, when the whole show is over, find a predicate in physics that was, in brute fact, coextensive with 'is a monetary exchange'. If physics is general—if the ontological biases of reductionism are true—then there must *be* such a predicate. But (a) to paraphrase a remark Professor Donald Davidson made in a slightly different context, nothing but brute enumeration could convince us of this brute coextensivity, and (b) there would seem to be no chance at all that the physical predicate employed in stating the coextensivity would be a physical kind term, and (c) there is still less chance that the coextension would be lawful (i.e., that it would hold not only for the nomologically possible world that turned out to be real, but for any nomologically possible world at all).[12]

[12] Oppenheim and Putnam (1958) argue that the social sciences probably *can* be reduced to physics assuming that the reduction proceeds via (individual) psychology. Thus, they remark, "in economics, if very weak assumptions are satisfied, it is possible to represent the way in which an individual orders his choices by means of an individual preference function. In terms of these functions, the economist attempts to explain group phenomena, such as the market, to account for collective consumer behavior, to solve the problems of welfare economics, etc." (p. 17). They seem not to have noticed, however, that even if such explanations can be carried through, they would not yield the kind of *predicate-by-predicate* reduction of economics to psychology that Oppenheim and Putnam's own account of the unity of science requires.

Suppose that the laws of economics hold because people have the attitudes, motives, goals, needs, strategies, etc., that they do. Then the fact that economics is the way it is can be explained by reference to the fact that people are the way that they are. But it doesn't begin to follow that the typical predicates of economics can be reduced to the typical predicates of psychology. Since bridge laws entail biconditionals, $P_1$ reduces to $P_2$ only if $P_1$ and $P_2$ are at least coextensive. But while the typical predicates of economics subsume (e.g.) monetary systems, cash flows, commodities, labor pools, amounts of capital invested, etc., the typical predicates of psychology subsume stimuli, responses, and mental states. Given the proprietary sense of 'reduction' at issue, to reduce economics to psychology would therefore involve a very great deal more than showing that the economic behavior of groups is determined by the psychology of the individuals that constitute them. In particular, it would involve showing that such notions as *commodity*, *labor pool*, etc., can be reconstructed in the vocabulary of stimuli, responses and mental states and that, moreover, the predicates which affect the reconstruction express psychological kinds (viz., occur in proper laws of psychology). I think it's fair to say that there is no reason at all to suppose that such reconstructions can be provided; prima facie there is every reason to think that they cannot.

I take it that the preceding discussion strongly suggests that economics is not reducible to physics in the special sense of reduction involved in claims for the unity of science. There is, I suspect, nothing peculiar about economics in this respect; the reasons why economics is unlikely to reduce to physics are paralleled by those which suggest that psychology is unlikely to reduce to neurology.

If psychology is reducible to neurology, then for every psychological kind predicate there is a coextensive neurological kind predicate, and the generalization which states this coextension is a law. Clearly, many psychologists believe something of the sort. There are departments of psychobiology or psychology and brain science in universities throughout the world whose very existence is an institutionalized gamble that such lawful coextensions can be found. Yet, as has been frequently remarked in recent discussions of materialism, there are good grounds for hedging these bets. There are no firm data for any but the grossest correspondence between types of psychological states and types of neurological states, and it is entirely possible that the nervous system of higher organisms characteristically achieves a given psychological end by a wide variety of neurological means. It is also possible that given neurological structures subserve many different psychological functions at different times, depending upon the character of the activities in which the organism is engaged.[13] In either event, the attempt to pair neurological structures with psychological functions could expect only limited success. Physiological psychologists of the stature of Karl Lashley have held this sort of view.

The present point is that the reductionist program in psychology is clearly *not* to be defended on ontological grounds. Even if (token) psychological events are (token) neurological events, it does not follow that the kind predicates of psychology are coextensive with the kind predicates of any other discipline (including physics). That is, the assumption that every psychological event is a physical event does not guarantee that physics (or, a fortiori, any other discipline more general than psychology) can provide an appropriate vocabulary for psychological theories. I emphasize this point because I am convinced that the make-or-break commitment of many physiological psychologists to the reductionist program stems precisely from having confused that program with (token) physicalism.

What I have been doubting is that there are neurological kinds coextensive with psychological kinds. What seems increasingly clear is that, even if there are such coextensions, they cannot be lawful. For it seems increas-

[13] This would be the case if higher organisms really are interestingly analogous to general purpose computers. Such machines exhibit no detailed structure-to-function correspondence over time; rather, the function subserved by a given structure may change from instant to instant depending upon the character of the program and of the computation being performed.

ingly likely that there are nomologically possible systems other than organisms (viz., automata) which satisfy the kind predicates of psychology but which satisfy no neurological predicates at all. Now, as Putnam has emphasized (1960a, b), if there are any such systems, then there must be vast numbers, since equivalent automata can, in principle, be made out of practically anything. If this observation is correct, then there can be no serious hope that the class of automata whose psychology is effectively identical to that of some organism can be described by *physical* kind predicates (though, of course, if token physicalism is true, that class can be picked out by some physical predicate or other). The upshot is that the classical formulation of the unity of science is at the mercy of progress in the field of computer simulation. This is, of course, simply to say that that formulation was too strong. The unity of science was intended to be an empirical hypothesis, defeasible by possible scientific findings. But no one had it in mind that it should be defeated by Newell, Shaw, and Simon.

I have thus far argued that psychological reductionism (the doctrine that every psychological natural kind is, or is coextensive with, a neurological natural kind) is not equivalent to, and cannot be inferred from, token physicalism (the doctrine that every psychological event is a neurological event). It may, however, be argued that one might as well take the doctrines to be equivalent since the only possible *evidence* one could have for token physicalism would also be evidence for reductionism: viz., that such evidence would have to consist in the discovery of type-to-type psychophysical correlations.

A moment's consideration shows, however, that this argument is not well taken. If type-to-type psychophysical correlations would be evidence for token physicalism, so would correlations of other specifiable kinds.

We have type-to-type correlations where, for every $n$-tuple of events that are of the same psychological kind, there is a correlated $n$-tuple of events that are of the same neurological kind.[14] Imagine a world in which such correlations are *not* forthcoming. What is found, instead, is that for every $n$-tuple of type identical psychological events, there is a spatiotemporally correlated $n$-tuple of type *distinct* neurological events. That is, every psychological event is paired with some neurological event or other, but psychological events of the same kind are sometimes paired with neurological events of different kinds. My present point is that such pairings would provide as much support for token physicalism as type-to-type pairings do *so long as we are able to show that the type distinct neurological events paired with a given kind of psychological event are identical in respect of whatever properties are relevant to type identification in psychology.* Suppose, for purposes of explication, that psychological events are type identified by reference to their behavioral consequences.[15] Then what is required of all the neurological events paired with a class of type homogeneous psychological events is only that they be identical in respect of their behavioral consequences. To put it briefly, type identical events do not, of course, have *all* their properties in common, and type distinct events must nevertheless be identical in *some* of their properties. The empirical confirmation of token physicalism does not depend on showing that the neurological counterparts of type identical psychological events are themselves type identical. What needs to be shown is just that they are identical in respect of those properties which determine what kind of *psychological* event a given event is.

*Could* we have evidence that an otherwise heterogeneous set of neurological events have those kinds of properties in common? Of course we could. The neurological theory might itself explain why an $n$-tuple of neurologically type distinct events are identical in their behavioral consequences, or, indeed, in respect of any of indefinitely many other such relational properties. And, if the neurological theory failed to do so, some science more basic than neurology might succeed.

My point in all this is, once again, not that correlations between type homogeneous psychological states and type heterogeneous neurological states would prove that token physicalism is true. It is only that such correlations might give us as much reason to be token physicalists as type-to-type correlations would. If this is correct, then epistemological arguments from token physicalism to reductionism must be wrong.

It seems to me (to put the point quite generally) that the classical construal of the unity of science has really badly misconstrued the *goal* of scientific reduction. The point of reduction is *not* primarily to find some natural kind predicate of physics coextensive with each kind predicate of a special science. It is, rather, to explicate the physical mechanisms whereby events conform to the laws of the special sciences. I have been arguing that there is no logical or epistemological reason why success in the second of these projects should require success in the first, and that the two are likely to come apart *in fact* wherever the physical mechanisms whereby events conform to a law of the special sciences are heterogeneous.

I take it that the discussion thus far shows that reductionism is probably too strong a construal of the unity of science; on the one hand, it is incompatible with probable results in the special sciences, and, on the other, it is more than we need to assume if what we primarily want, from an ontological point of view, is just to be good token physicalists. In what follows, I shall try to sketch a liberalized version of the relation between physics and

[14] To rule out degenerate cases, we assume that $n$ is large enough to yield correlations

[15] I don't think there is any chance at all that this is true. What is more likely is that type identification for psychological states can be carried out in terms of the 'total states' of an abstract automaton which models the organism whose states they are.

the special sciences which seems to me to be just strong enough in these respects. I shall then give a couple of independent reasons for supposing that the revised doctrine may be the right one.

The problem all along has been that there is an open empirical possibility that what corresponds to the kind predicates of a reduced science may be a heterogeneous and unsystematic disjunction of predicates in the reducing science. We do not want the unity of science to be prejudiced by this possibility. Suppose, then, that we allow that bridge statements may be of this form,

(4) $Sx \rightleftharpoons P_1x \vee P_2x \vee \ldots \vee P_nx$

where $P_1 \vee P_2 \vee \ldots \vee P_n$ is *not* a kind predicate in the reducing science. I take it that this is tantamount to allowing that at least some 'bridge laws' may, in fact, not turn out to be laws, since I take it that a necessary condition on a universal generalization being lawlike is that the predicates which constitute its antecedent and consequent should be kind predicates. I am thus assuming that it is enough, for purposes of the unity of science, that every law of the special sciences should be reducible to physics by bridge statements which express true empirical generalizations. Bearing in mind that bridge statements are to be construed as species of identity statements, formula (4) will be read as something like 'every event which consists of $x$'s satisfying $S$ is identical with some event which consists of $x$'s satisfying some or other predicate belonging to the disjunction $P_1 \vee P_2 \vee \ldots \vee P_n$'.

Now, in cases of reduction where what corresponds to formula (2) is not a law, what corresponds to formula (3) will not be either, and for the

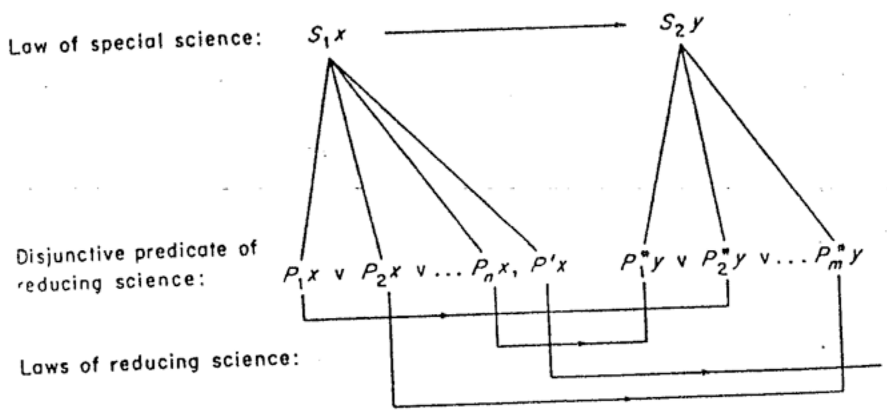


**Figure I-1** Schematic representation of the proposed relation between the reduced and the reducing science on a revised account of the unity of science. If any $S_1$ events are of the type $P'$, they will be exceptions to the law $S_1x \to S_2y$. See text.

same reason: viz., the predicates appearing in the antecedent and consequent will, by hypothesis, not be kind predicates. Rather, what we will have is something that looks like Figure I-1. That is, the antecedent and consequent of the reduced law will each be connected with a disjunction of predicates in the reducing science. Suppose, for the moment, that the reduced law is exceptionless, viz., that no $S_1$ events satisfy $P'$. Then there will be laws of the reducing science which connect the satisfaction of *each* member of the disjunction associated with the antecedent of the reduced law with the satisfaction of some member of the disjunction associated with the consequent of the reduced law. That is, if $S_1x \to S_2y$ is exceptionless, then there must be some proper law of the reducing science which either states or entails that $P_1x \to P^*$ for some $P^*$, and similarly for $P_2x$ through $P_nx$. Since there must be such laws, and since each of them is a 'proper' law in the sense in which we have been using that term, it follows that each disjunct of $P_1 \vee P_2 \vee \ldots \vee P_n$ is a kind predicate, as is each disjunct of $P^*_1 \vee P^*_2 \vee \ldots \vee P^*_n$.

This, however, is where push comes to shove. For it might be argued that if each disjunct of the $P$ disjunction is lawfully connected to some disjunct of the $P^*$ disjunction, then it follows that formula (5) is itself a law.

(5) $P_1x \vee P_2x \vee \ldots \vee P_nx \to P^*_1y \vee P^*_2y \vee \ldots \vee P^*_ny$

The point would be that the schema in Figure I-1 implies $P_1x \to P^*_2y$, $P_2x \to P^*_my$, etc., and the argument from a premise of the form $(P \supset R)$ *and* $(Q \supset S)$ to a conclusion of the form $(P \vee Q) \supset (R \vee S)$ is valid.

What I am inclined to say about this is that it just shows that 'it's a law that __________' defines a nontruth functional context (or, equivalently for these purposes, that not all truth functions of kind predicates are themselves kind predicates); in particular, that one may not argue from: 'it's a law that $P$ brings about $R$' and 'it's a law that $Q$ brings about $S$' to 'it's a law that $P$ or $Q$ brings about R or $S$'. (Though, of course, the argument from those premises to '$P$ or $Q$ brings about $R$ or $S$' *simpliciter* is fine.) I think, for example, that it is a law that the irradiation of green plants by sunlight causes carbohydrate synthesis, and I think that it is a law that friction causes heat, but I do not think that it is a law that (either the irradiation of green plants by sunlight or friction) causes (either carbohydrate synthesis or heat). Correspondingly, I doubt that 'is either carbohydrate synthesis or heat' is plausibly taken to be a kind predicate.

It is not strictly mandatory that one should agree with all this, but one denies it at a price. In particular, if one allows the full range of truth-functional arguments inside the context 'it's a law that __________', then one gives up the possibility of identifying the kind predicates of a science with the ones which constitute the antecedents or consequents of its proper laws. (Thus formula (5) would be a proper law of physics which fails to satisfy that condition.) One thus inherits the need for an alternative construal of the notion of a kind, and I don't know what that alternative would be like.

The upshot seems to be this. If we do not require that bridge statements must be laws, then either some of the generalizations to which the laws of special sciences reduce are not themselves lawlike, or some laws are not formulable in terms of kinds. Whichever way one takes formula (5) the important point is that the relation between sciences proposed by Figure I-1 is weaker than what standard reductionism requires. In particular, it does not imply a correspondence between the kind predicates of the reduced and the reducing science. Yet it does imply physicalism given the same assumption that makes standard reductionism physicalistic: viz., that bridge statements express token event identities. But these are precisely the properties that we wanted a revised account of the unity of science to exhibit.

I now want to give two further reasons for thinking that this construal of the unity of science is right. First, it allows us to see how the laws of the special sciences could reasonably have exceptions, and, second, it allows us to see why there are special sciences at all. These points in turn.

Consider, again, the model of reduction implicit in formulae (2) and (3). I assume that the laws of basic science are strictly exceptionless, and I assume that it is common knowledge that the laws of the special sciences are not. But now we have a dilemma to face. Since '→' expresses a relation (or relations) which must be transitive, formula (1) can have exceptions only if the bridge laws do. But if the bridge laws have exceptions, reductionism loses its ontological bite, since we can no longer say that every event which consists of the satisfaction of an *S*-predicate consists of the satisfaction of a *P*-predicate. In short, given the reductionist model, we cannot consistently assume that the bridge laws and the basic laws are exceptionless while assuming that the special laws are not. But we cannot accept the violation of the bridge laws unless we are willing to vitiate the ontological claim that is the main point of the reductionist program.

We can get out of this (*salve* the reductionist model) in one of two ways. We can give up the claim that the special laws have exceptions or we can give up the claim that the basic laws are exceptionless. I suggest that both alternatives are undesirable—the first because it flies in the face of fact. There is just no chance at all that the true, counterfactual supporting generalizations of, say, psychology, will turn out to hold in strictly each and every condition where their antecedents are satisfied. Even when the spirit is willing the flesh is often weak. There are always going to be behavioral lapses which are physiologically explicable but which are uninteresting from the point of view of psychological theory. But the second alternative is not much better. It may, after all, turn out that the laws of basic science have exceptions. But the question arises whether one wants the unity of science to depend on the assumption that they do.

...d in Figure I-1, however, everything works out satisfactorily. A nomologically sufficient condition for an exception to $S_1x \rightarrow S_2y$ is that the bridge statements should identify some occurrence of the satisfaction of $S_1$ with an occurrence of the satisfaction of a *P*-predicate which is not itself lawfully connected to the satisfaction of any $P^*$-predicate (i.e., suppose $S_1$ is connected to $P'$ such that there is no law which connects $P'$ to any predicate which bridge statements associate with $S_2$. Then any instantiation of $S_1$ which is contingently identical to an instantiation of $P'$ will be an event which constitutes an exception to $S_1x \rightarrow S_2y$). Notice that, in this case, we need assume no exceptions to the laws of the *reducing* science since, by hypothesis, formula (5) is not a law.

In fact, strictly speaking, formula (5) has no status in the reduction at all. It is simply what one gets when one universally quantifies a formula whose antecedent is the physical disjunction corresponding to $S_1$ and whose consequent is the physical disjunction corresponding to $S_2$. As such, it will be true when $S_1x \rightarrow S_2y$ is exceptionless and false otherwise. What does the work of expressing the physical mechanisms whereby *n*-tuples of events conform, or fail to conform, to $S_1x \rightarrow S_2y$ is not formula (5) but the laws which severally relate elements of the disjunction $P_1 \vee P_2 \vee \ldots \vee P_n$ to elements of the disjunction $P^*_1 \vee P^*_2 \vee \ldots \vee P^*_m$. Where there *is* a law which relates an event that satisfies one of the *P* disjuncts to an event which satisfies one of the $P^*$ disjuncts, the pair of events so related conforms to $S_1x \rightarrow S_2y$. When an event which satisfies a *P*-predicate is not related by law to an event which satisfies a $P^*$-predicate, that event will constitute an exception to $S_1x \rightarrow S_2y$. The point is that none of the laws which effect these several connections need themselves have exceptions in order that $S_1x \rightarrow S_2y$ should do so.

To put this discussion less technically: We could, if we liked, *require* the taxonomies of the special sciences to correspond to the taxonomy of physics by insisting upon distinctions between the kinds postulated by the former whenever they turn out to correspond to distinct kinds in the latter. This would *make* the laws of the special sciences exceptionless if the laws of basic science are. But it would also likely loose us precisely the generalizations which we want the special sciences to express. (If economics were to posit as many *kinds* of monetary systems as there are physical realizations of monetary systems, then the generalizations of economics *would* be exceptionless. But, presumably, only vacuously so, since there would be no generalizations left for economists to state. Gresham's law, for example, would have to be formulated as a vast, open disjunction about what happens in monetary $\text{system}_1$ or monetary $\text{system}_n$ under conditions which would themselves defy uniform characterization. We would not be able to say what happens in monetary systems *tout court* since, by hypothesis, 'is a monetary system' corresponds to no kind predicate of physics.)

In fact, what we do is precisely the reverse. We allow the generalizations of the special sciences to *have* exceptions, thus preserving the kinds to which the generalizations apply. But since we know that the *physical* descriptions of the members of these kinds may be quite heterogeneous, and since we know that the physical mechanisms which connect the satisfaction of the antecedents of such generalizations to the satisfaction of their consequents may be equally diverse, we expect both that there will be exceptions to the generalizations and that these will be 'explained away' at the level of the reducing science. This is one of the respects in which physics really is assumed to be bedrock science; exceptions to *its* generalizations (if there are any) had better be random, because there is nowhere 'further down' to go in explaining the mechanism whereby the exceptions occur.

This brings us to why there are special sciences at all. Reductionism, as we remarked at the outset, flies in the face of the facts about the scientific institution: the existence of a vast and interleaved conglomerate of special scientific disciplines which often appear to proceed with only the most casual acknowledgment of the constraint that their theories must turn out to be physics 'in the long run'. I mean that the acceptance of this constraint often plays little or no role in the practical validation of theories. Why is this so? Presumably, the reductionist answer must be *entirely* epistemological. If only physical particles weren't so small (if only brains were on the *out*side, where one can get a look at them), *then* we would do physics instead of paleontology (neurology instead of psychology, psychology instead of economics, and so on down). There is an epistemological reply: viz., that even if brains were out where they could be looked *at*, we wouldn't, as things now stand, know what to look *for*. We lack the appropriate theoretical apparatus for the psychological taxonomy of neurological events.

*If* it turns out that the functional decomposition of the nervous system corresponds precisely to its neurological (anatomical, biochemical, physical) decomposition, then there are only epistemological reasons for studying the former instead of the latter. But suppose that there is no such correspondence? Suppose the functional organization of the nervous system cross-cuts its neurological organization. Then the existence of psychology depends not on the fact that neurons are so depressingly small, but rather on the fact that neurology does not posit the kinds that psychology requires.

I am suggesting, roughly, that there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together: not all the kinds (not all the classes of things and events about which there are important, counterfactual supporting generalizations to make) are, or correspond to, physical kinds. A way of stating the classical reductionist view is that things which belong to different physical kinds ipso facto can have none of their projectable descriptions in common[16]: that if $x$ and $y$ differ in those descriptions by virtue of which they fall under the proper laws of physics, they must differ in those descriptions by virtue of which they fall under any laws at all. But why should we believe that this is so? Any pair of entities, however different their physical structure, must nevertheless converge in indefinitely many of their properties. Why should there not be, among those convergent properties, some whose lawful interrelations support the generalizations of the special sciences? Why, in short, should not the kind predicates of the special sciences *cross-classify* the physical natural kinds?[17]

Physics develops the taxonomy of its subject matter which best suits its purposes: the formulation of exceptionless laws which are basic in the several senses discussed above. But this is not the only taxonomy which may be required if the purposes of science in general are to be served: e.g., if we are to state such true, counterfactual supporting generalizations as there are to state. So there are special sciences, with their specialized taxonomies, in the business of stating some of these generalizations. If science is to be unified, then all such taxonomies must apply *to the same things*. If physics is to be basic science, then each of these things had better be a physical thing. But it is not further required that the taxonomies which the special sciences employ must themselves reduce to the taxonomy of physics. It is not required, and it is probably not true.

Try as they may, many philosophers find it hard to take literally the things that nonphilosophers say. Since verificationism became unfashionable, most philosophers have conceded—some have even insisted—that the claims of the laity are often true when they are construed correctly. But the correct construal is frequently far to seek, and almost always proves remarkably different from what the laity had thought it had in mind. Thus, for a while, philosophers taught that talking about tables and chairs is an elliptical and misleading way of referring to the states of one's visual field and warned that the foundations of inductive inference would surely crumble unless physical objects turned out to be 'constructs' out of phenomena logically homogeneous with afterimages. In the event, however, 'physical object talk' was found to require considerably less analysis than had been supposed. Tables and chairs proved to be not at all like afterimages, and the practice of inductive inference survived.

But while reductionism is now widely deplored in epistemology proper, it lingers in philosophical discussions of 'theoretical constructs' in the sciences.

[16] For the notion of projectability, see Goodman (1965). All projectible predicates are kind predicates, though not, presumably, vice versa.

[17] As, by the way, the predicates of natural languages quite certainly do. (For discussion, see Chomsky, 1965.)

To assert that the taxonomies employed by the special sciences cross-classify physical kinds is to deny that the special sciences, together with physics, constitute a hierarchy. To deny that the sciences constitute a hierarchy is to deny precisely what I take the classical doctrine of the unity of science to assert insofar as it asserts anything more than token physicalism.

Psychological theories, in particular, have struck many philosophers as apt for dehypostatization, and the warnings that the alternative to reduction is a ruinous skepticism have an all too familiar ring. It has, however, been the burden of these introductory remarks that the arguments for the behavioral or physiological reduction of psychological theories are not, after all, very persuasive. The results of taking psychological theories literally and seeing what they suggest that mental processes are like might, in fact, prove interesting. I propose, in what follows, to do just that.

hypostasis: essential substance.

# FIRST APPROXIMATIONS

14

*I'm the only President you've got.*
LYNDON B. JOHNSON

The main argument of this book runs as follows:

1. The only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational.
2. Computation presupposes a medium of computation: a representational system.
3. Remotely plausible theories are better than no theories at all.
4. We are thus provisionally committed to attributing a representational system to organisms. 'Provisionally committed' means: committed insofar as we attribute cognitive processes to organisms and insofar as we take seriously such theories of these processes as are currently available.
5. It is a reasonable research *goal* to try to characterize the representational system to which we thus find ourselves provisionally committed.
6. It is a reasonable research *strategy* to try to infer this characterization from the details of such psychological theories as seem likely to prove true.
7. This strategy may actually work: It is possible to exhibit specimen inferences along the lines of item 6 which, if not precisely apodictic, have at least an air of prima facie plausibility.

The epistemic status of these points is pretty various. I take it, for example, that item 3 is a self-evident truth and therefore requires no justification beyond an appeal to right reason. I take it that item 4 follows from items 1–3. Items 5–7, on the other hand, need to be justified *in practice*. What must be shown is that it is, in fact, productive to conduct psychological research along the lines they recommend. Much of the material in later chapters of this book will be concerned to show precisely that. Hence, the discussion will become more intimately involved with empirical findings, and with their interpretations, as we go along.