

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/16087638>

# Mind-brain reduction: New light from the philosophy of science

Article in *Neuroscience* · June 1982

DOI: 10.1016/0306-4522(82)91117-4 · Source: PubMed

---

CITATIONS

31

---

READS

212

1 author:



[Pat Churchland](#)

University of California, San Diego

128 PUBLICATIONS 7,294 CITATIONS

[SEE PROFILE](#)

## COMMENTARY

# MIND-BRAIN REDUCTION: NEW LIGHT FROM THE PHILOSOPHY OF SCIENCE

P. S. CHURCHLAND

University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

**Abstract**—The discussion of the reduction of mental states to brain states is placed in the broader context of reduction in other scientific disciplines such as chemistry, physics and biology. This is important in achieving a perspective on what sort of business reduction is and in seeing that reductions are primarily transitions between theories and only derivatively relations between phenomena. It also reveals that though reducing theories sometimes absorb the old theory as largely correct, more often the old theory is substantially modified and revised and sometimes it is replaced outright. How much of the old theory survives in the reducing theory depends on its empirical integrity and whether its basic categories are empirically sound. The reduction of psychology to neuroscience is considered in this light and it is suggested that psychology may be substantially revised or even replaced by a reduction to neuroscience.

A singularly productive development in philosophy in the last several decades has been the joining and intermixing of philosophy of science and philosophy of mind. The standard problems concerning the mind include the following: what is the relation between mind and the brain; are mental states identical to brain states; what is the status of introspective reports—do they, for example, contain incontrovertible information about the nature of the subject's mind-brain; what is special about the subjective point of view; what would a reduction of mental states to brain states involve? Broadly speaking, the reason this development was so significant is that many insights garnered from the study of how sciences in fact develop and evolve, together with insights about the nature of reduction and explanation, were found to be applicable with stunning success to problems in the philosophy of mind. Many of the above questions have been found to have instructive analogues in the history of other sciences and, as well as providing much prized insight, this has had the beneficent effect of challenging the assumption that questions in the philosophy of mind are utterly unique and without precedent, that they can or should be addressed in isolation from the rest of physics, biology, neuroscience and psychology, or that they are somehow not, after all, scientific questions. Part and parcel of this shift in perspective is the broader view that philosophy at its best and properly conceived is continuous with science, differing from the specific scientific disciplines mainly in its scope and generality, but not in its ultimately empirical and testable nature. My aim in this paper is to render accessible to non-philosophers some of the insights resulting from the inter-

animation of philosophy of science and philosophy of mind, especially as they bear upon the question of reduction.

### INTERTHEORETIC REDUCTION

'Reductionism' has unfortunately become something of a 'boo' word in some quarters, apt to connote a scorn for humankind or a disdain for moral conceptions based on notions of responsibility and deliberation. Doubtless, such recently acquired connotations are owed in part to the fact that Skinner vociferously proclaimed himself a reductionist and his favoured hypothesis of what form the reduction of folk psychology would take was not only empirically far-fetched, but was frequently accompanied with bombastic prophecies designed to shock and outrage. Additionally, it is a stock rhetorical tactic practised by hide-bound devotees of the 'non-physical mind' to cast reductionists in the villain's role (see, for example, Popper & Eccles<sup>16</sup>). Chilling connotations aside, reduction is essentially just a relation between theories and if one phenomenon is said to be reducible to another, this is in virtue of the theory which describes that phenomenon reducing to a more fundamental theory. Reduction in the physical and biological sciences is typically considered a good thing, insofar as it brings about a unification of explanation and coherence in understanding. Theoretical and explanatory integration is the goal of reductive strategies in the sciences; theoretical balkanization implies that pieces of the puzzle are still missing.

The pressing question then is: what is the inter-theoretic relation? When is an old theory successfully

---

Abbreviation: STR, Special Theory of Relativity.

reduced by a new one? The fast answer, to be elaborated anon, is this: a reduction obtains when the *terms* of the old theory can be mapped onto (i.e. paired with) expressions in the new theory, in such a way that the basic *laws* of the old theory are thereby mapped onto sentences in the new theory, where these sentences are *logical consequences* of the basic laws of the new theory. In effect, an image of the old, reduced theory is generable in the new, reducing theory, which means that the new theory can explain what the old theory explained and can explain why the old theory worked as well as it did, and of course the new theory typically explains much where the old theory was buffaloed.

The fast answer provides a general framework for the discussion, but clarification of certain crucial notions is needed, especially for the all-important mapping relation between terms and between sentences of the old and new theories. In the relatively simple and smooth cases of reduction, the mapping of terms will yield identity statements, e.g. the temperature of a gas = mean kinetic energy of its constituent molecules. As for the laws of the old theory, they will find closely resembling sentences in the language of the new theory, where these latter will be logically derived from the laws of the new theory, e.g. the classical gas law,  $PV = \mu RT^*$ , closely resembles the sentence  $PV = (2N/3)(mv^2/2)^\dagger$  which, together with certain assumptions, such as the perfect elasticity of molecules, is derivable from the laws of mechanics. While the reduction of the classical gas law to statistical mechanics is an elegant and perhaps the best beloved case of reduction, its elegance and simplicity should not beguile us into supposing it either typical of reduction in the sciences, or even the archetype beckoning our aspirations. The evolution of theories is a natural phenomenon and reduction-on-the-hoof rarely exhibits textbook simplicity. The history of science reveals that the mapping relations between terms are seldom unproblematic identity statements, and the construction of an image of the old theory in the new to devise a derivation typically involves varying degrees of correlation and revision of the old laws. This point is absolutely crucial in considering the reduction of psychology to neuroscience, since all too frequently it is assumed that reduction requires identity statements, and that if the identity of mental states—as understood within current psychological theory—and brain states is implausible, then the reductive strategy is doomed. For example, it is often supposed that either mental states (as now conceived) are identical to brain states (as now conceived) or mind/brain dualism is true. No such conclusion fol-

lows and the dilemma is misbegotten. The success of reductive strategies does not require that mapping relations yield identity statements, and indeed, reductive strategies may be cramped and impeded if neuroscientists suppose they are obliged to confine their search for new theory to one which will permit preservation of the old theory. Unless we are uncommonly lucky, current conceptions of mental states and processes will have to be thoroughly revised and corrected at the behest of new discoveries, and to the extent that they have to be corrected and revised, the likelihood of getting identity statements fades.

Recent work of philosophers and historians of science has provided new understanding of reduction in its many guises (see, especially, Hooker,<sup>11</sup> Causey,<sup>1</sup> Feyerabend,<sup>9</sup> Schaffner,<sup>18</sup> Wimsatt<sup>22</sup>) and I want now to introduce some examples from that research in illustration.

Special Theory of Relativity (henceforth, STR) succeeds Classical or Newtonian Mechanics and superficially it may appear that a mapping of the terms in Classical Mechanics to terms in STR yields identity statements to the effect that mass denoted in Classical Mechanics ( $Mass_C$ ) is identical to mass denoted in STR ( $Mass_{STR}$ ). Closer scrutiny soon reveals this cannot be so and that the putative identities are no such thing.  $Mass_{CM}$ , for example, is an *intrinsic feature* of an entity, but  $Mass_{STR}$  is a *relation* holding between an entity and countless reference frames.<sup>2</sup> Properties intrinsic to entities are not relations between entities, no matter how you cut it. Here, then, is an instance of theory reduction where an entire framework of concepts and laws is reconfigured and identity statements play no role. If STR is right, there is no such thing as  $Mass_{CM}$  and conversely, if Classical Mechanics is right, there is no such thing as  $Mass_{STR}$ ; hence, there is nothing described in Classical Mechanics for  $Mass_{STR}$  to be identical to. Evidently, in this instance of theory reduction the mapping of terms functions more like a rule for the replacement of old terms with new, superficial similarity notwithstanding.

Moreover, it is important to notice that the laws of Classical Mechanics cannot be derived from STR, though a recognizably similar but reconfigured analogue of those laws can be. The laws of Classical Mechanics cannot, undoctored, be derived from STR for two reasons: (1) the terms of those laws are radically different (properties versus relations) and (2) even when the terms of the old are replaced by terms of the new theory, the old laws thus suitably doctored apply only within certain limits and this must be factored in as a (false) assumption that such limits obtain.

Once done, the image of Classical Mechanics can be generated in STR and we can see why Classical Mechanics worked as well as it did, why it failed in certain places and why it was dumbstruck in others.

What emerges from a study of this case is that theory reduction may take the form not of preserving the old theory under the wing of the new, but of

\* Where  $\mu$  is the amount of gas in fractions of a mol, and  $R$  is the gas constant.

† Where  $N$  is the number of molecules of gas in the volume,  $m$  is individual molecular mass, and  $v$  is average molecular velocity.

showing that the old theory was systematically misconceived and illusory, that its key concepts were essentially misdirected and that the reduced theory is actually rejected and replaced rather than retained. Other case studies reveal a similar pattern and it appears that retention of a theory through reduction is not an all-or-nothing affair, but rather a matter of degree, with some cases showing substantial retention of the old theory, others showing substantial replacement. Virtually every case of reduction involves some correction of the old theory before it (its analogue) can be derived from the new and this betokens reconfiguration of the terms which in turn betokens the evanescence of identities. Determining when the reconfiguration of terms is so minor that identity statements can be claimed is like clawing at the air, and certainly no useful formal criteria are available. Nor are such criteria likely to be forthcoming, since pragmatic and social considerations figure crucially; the whim of the central investigators, the degree to which confusion will result from retention of old terms, the desire to preserve or to break with past habits of thought, the related opportunities for publicising the theory and for cadging grants and graduate students, all enter into decisions concerning whether to claim identities and therewith retention, or to make the more radical claim of replacement. In fact, I do not think it matters much that we establish criteria for determining when retention of the old theory is sufficient to claim identity statements. What finally does matter is whether the new theory is superior to the incumbent; that is, whether it is explanatorily more powerful or not, whether it is consonant or antithetical to scientific unity and whether it expands our understanding.

Several other examples should be touched on at least briefly in order to round out the case that reduction does not require identities. At the replacement end of the continuum is the reduction of phlogiston theory to oxygen theory, and the reduction of the caloric theory of heat to the energy theory, where in both cases the reducing theories diverged so profoundly from the reduced, that these latter were considered thoroughly muddled and mistaken, not just in the details but in the rudiments. There is no such thing as caloric fluid and notice that we do not try to salvage the old concept by seeking to jury-rig an identity of caloric fluid and kinetic energy. Also noteworthy here is the reduction of Ptolemaic theory to Newtonian theory (via Copernican theory), for in this case at least one *observable* phenomenon of the old theory is denied existence in the new theory. This deserves a small aside.

In Ptolemaic theory, the earth was thought to be enclosed within a celestial sphere, rather like a pea enclosed within the walls of a gigantic basketball, save of course that the celestial sphere was composed of grand and precious (i.e. celestial) stuff. On this theory, the sphere was considered as observable as the sun or the stars and it could be seen to turn daily around the

inert and centrally stationed earth. Copernicus brought the sphere to rest and made it much larger. But there was no place for such a sphere in Newton's infinite universe and the triumph of this new view completely undercut the idea that when we look at the night sky we are observing the inside walls of an enclosing sphere. As this example illustrates, when a new theory invites radical reconceptualization, even our observations may be reconfigured. There is no crystal sphere to observe, though for much of man's history it was considered unproblematically and directly observable.

This is an important point, especially since the old positivist doctrine that there is a domain of theory-free observation has come to be discredited in the last two decades.<sup>2,8,11</sup> Observations are not, alas, the neutral touchstone permitting impartial arbitration between competing theories. Our observations of the world do not come innocent of beliefs about what the world is like, about what things can be observed and how they appear; rather, our observations are highly interpreted, processed and conceptualized, and the concepts involved form the nodes of a theoretical network which may itself be empirically false. This will loom increasingly significant when we contemplate the possibility that our psychological categories for thinking about ourselves might be mistaken—as mistaken as such categories as caloric fluid, or the starry sphere of the heavens.

Cases which fall more on the retention end of the spectrum would include the reduction of physical optics to electromagnetic theory, as well as my first mentioned case, the reduction of classical thermodynamics to statistical mechanics. Not even here, however, is there perfect retention, for in both cases there is considerable correction of the old theory before it can be derived from the new. For example, entropy, a single concept of classical thermodynamics, has no counterpart in statistical mechanics since it is not to be identified with Boltzman entropy, or with either fine grained or coarse grained Gibbs entropies.<sup>14</sup> Sundry cases fall between the two extremes of theoretical retention and theoretical replacement. Rigor in plotting just where in the spectrum the cases fall, or just how retentive a reduction is, remains elusive, nor is there much profit to be gained by aiming for it. For one thing, often on their way to reductive concert, theories co-evolve as each informs and influences the other. In such cases, the original theories undergo much modification and transformation and it is the modified and embellished progeny of the original theories which consummate the reduction. This is well exemplified by transmission genetics and molecular genetics which have already undergone considerable co-evolution<sup>11,12</sup> each providing results and problems for the other, with consequent modifications, corrections and extensions being made to both theories. As things stand, the genes of transmission genetics are not identifiable with the genes of molecular genetics, though future modification of both theories might

result in the identification. Alternatively, if it does not, the forecast reduction may be more replacing than retaining in character. Co-evolution of theories is an incessant and inveterate feature of theory progression, and it belies the textbook characterization of theories as stable and static entities, devoid of a dynamic. This too will loom significant in considering the relations between neuroscience and psychology.

### REDUCTION TO NEUROSCIENCE

The points which I have been concerned to emphasize so far are these: (1) reduction is, at bottom, a relation between theories; (2) reduction may retain much of the old theory, or may replace wholesale the old theory, or may be somewhere in between. Only reduction toward the retention end of the spectrum provides identity statements specifying phenomena described in the reduced theory as the same phenomena described in the reducing theory; (3) theories co-evolve. This is essential to entertaining questions concerning the reduction of mental states to brain states.

Accordingly, the first matter to address is this: when the reduction of mental states to brain states is considered, what is the *theory* describing mental states which is up for reduction to neuroscientific theory? The customary answer since the mid-sixties points to that theory, call it 'folk psychology', in virtue of which we standardly and routinely observe and explain the behaviour of persons.<sup>2,19</sup> Our commonsense, everyday conceptual network, which we exploit in explaining why a person did what he did, is a theory of sorts, embedded in the language learned at mother's knee and so taken for granted in observation of oneself and others that it is scarce recognized for the theory it is. This theory is really just a body of collected lore by means of which we ascribe beliefs, desires, perceptions, sensations, consciousness, etc., to others and to ourselves. Beliefs and desires form the crux of the explanation of behaviour and hence sometimes the theory is called 'belief-desire psychology'<sup>20</sup> and sometimes 'folk psychology'.<sup>5,7,10</sup> Much of the time our explanation and prediction of behaviour is fairly humdrum, as when one predicts the busdriver will stop at the Cordova Street bus stop and explains why he did in terms of his hearing the bell ring, his belief that there was a passenger who wanted to get off and his desire to let the passenger off. Failures to do what we expect are often also explicable: the busdriver failed to stop because he did not hear the bell and so did not have the belief that a passenger wished to get off. Generalizations concerning beliefs, desires, perceptions, and deliberate action are implicit in much of the explaining we do. Our understanding of mental states is crucially dependent on the conceptual network which describes and interconnects them.

Because this body of collected lore, this 'folk psychology', is inherited as part of the commonsense conception of things, rather than invented with careful

deliberation, it has sometimes seemed odd to call it a theory. Yet to assign or withhold the status of 'theory' according to origins is not very useful. Other aboriginal conceptual networks, long part of commonsense but now seen to be false, are typically considered to be theories—their falsity presumably ensuring that status. So origin alone cannot be the deciding factor here. Of course it may be thought that folk psychology, unlike say, animism or geocentrism, is not false. Perhaps so, but that after all is an empirical question, and whether folk psychology really is true is anyhow irrelevant to whether it should be considered a theory. Moreover, in its heyday, *any commonsense* theory seems obviously true, which is how it earns the honorific 'common sense'. If folk psychology seems obviously true, that is a consequence of its superincumbent familiarity, not of its having survived stern empirical scrutiny.

Apart from the fact that folk psychology has intrinsic features which make it appropriately conceived as a theory, it is extremely useful so to conceive it. For then questions can be asked about how good a theory it is, about whether it has explanatory failures and how serious these are, about how it has developed or failed to develop, about how and whether it might be expanded and developed, about how it might make contact with other scientific disciplines and how it might reduce to neuroscience. Once seen as a theory, folk psychology can be held at arm's length and considered the way we consider any scientific theory, not simply taken for granted because it is all we have. At least, it makes sense to wonder whether the fundamental categories of folk psychology might be hopelessly amiss and whether folk psychology might eventually go the route of alchemy and animism. Accordingly, when contemplating the reduction of psychology to neuroscience, we should not prejudge these questions by assuming that the reduction will be retentive in nature, providing identifications of mental states—as currently conceived—with brain states. The reduction may require a lengthy period of co-evolution and the psychology which is finally reduced to neuroscience may bear only remote resemblance to folk psychology as it now stands.

Having liberated ourselves from the idea that a reduction can be achieved only if an identity of folk psychological terms and neuroscientific terms is nailed down, the next step is to try to get a fix on how the reduction might go, that is, whether it is likely to be largely retentive in style, or whether most of folk psychology will be replaced. One way to begin this task is to take a dispassionate look at folk psychology as an empirical theory of how the mind-brain works and to see how well or ill it fares. If it seems to have features that mark it as a dud in its empirical role, then replacement may be on the cards. Since the empirical worthiness and resiliency of folk psychology often seems patently obvious and beyond serious dispute, a balanced assessment may be aided by cataloguing its flaws, faults, and shortcomings. In what fol-

lows I shall emphasize, briefly, several features which bestir misgivings about the adequacy of folk psychology and which prompt the suspicion that folk psychology, as we know and love it, will eventually find its place beside vitalism, animism, geocentrism and other curios of commonsense science.

Although folk psychology serves decently well in much of the everyday business of explaining behaviour (as in the case of the compliant busdriver), it falters badly outside this humdrum domain. Among the things it is powerless to explain include such commonplace phenomena as the nature and function of sleep and dreaming, how information is stored and retrieved, how learning takes place, how language is acquired, how growing and learning are related, the nature and cause of mental dysfunction, and so on and on. This list is long, and the recalcitrant phenomena are neither *recherché* nor unusual. To be sure, much research done by psychologists is trained on these problems, but it seems to me that, as often as not, either the results lead to answers which are wholly physiological in nature, or they lead to the exposure of new problems and difficulties with folk psychology rather than to a smooth development of it. As to the general matter of progress, folk psychology can boast of virtually no significant correction or extension since oldest recorded texts. Plato and Hippocrates explained human behaviour in essentially the same way we do now and were puzzled and stupefied by the phenomena which puzzle and stupefy us. In contrast to the monumental progress in, say, biology or physics, folk psychology cannot claim that in the last 2000 years it has become measurably richer or more encompassing, or that it has deepened our understanding of how the mind-brain works. Seen in this light, folk psychology looks like a stagnant research program, doomed to frustrating ineptitude by its conceptual framework—by its misconceptions. By contrast, its main competitor, physiological psychology, looks robust and flourishing and not only has an ongoing program and bountiful results, but also makes contact with the rest of science, a feature which augers well for unity of explanation and understanding.<sup>3</sup> To be fair, it must be mentioned that folk psychology does indeed have its champions, foremost among which are perhaps the cognitive psychologists. The crux of the new movement, cognitive science, is that folk psychology is at bottom sound enough and correct enough to sustain real development into a scientific psychology. Whether this strategy will produce much of significance remains to be seen, but it is noteworthy that cognitive psychologists frequently aver that determining the cognitive functions of the mind-brain can be carried out independently of knowing anything about neurophysiology. (I have criticized this view extensively in P. S. Churchland.<sup>5,6</sup>)

In addition to its distinctly pinched explanatory province, folk psychology has other problems which might best be called methodological, and which will, I suspect, dog the footsteps of the cognitive movement.

Earlier, I mentioned that the heart of folk psychology is the belief–desire axis and now it must be emphasized that the signal fact about beliefs and desires is that they are *sentential attitudes*. What this means is that beliefs are always beliefs *that p*, where *p* is some sentence or other. Thus A believes that Philby was a spy, or B believes that everyone on the jury suspects he is guilty, etc. Similarly, for desires. The states are identified by means of a sentence, sometimes called the ‘content sentence’ because it gives the content of the states. And so it goes for thinking that *p*, perceiving that *p*, fearing, expecting, hoping, wondering, etc., that *p*. The second point is this: transitions between states are taken to be essentially logical operations on the content sentences. On this view, it is not so much the causal relations between mind-brain states *physiologically* described that matter in explanation, but the logical relations between mind-brain states *sententially* described that matter. Folk psychology then, is basically a sentential psychology.

There are several reasons for suspecting this bodes ill for folk psychology and, indeed, for any psychological theory which is devoted to sentential attitudes. To begin with, it implies that the brain’s cognitive operations really are operations on sentences, and while some cognitive psychologists have recognized this and have been willing stalwartly to live with it,<sup>10</sup> it is not easy. A number of considerations militate against this and I shall mention two here. (For more, see Churchland,<sup>5</sup> Dennett,<sup>7</sup> Stich.<sup>21</sup>) First, it causes no end of trouble in dealing with the cognitive operations and intelligent behaviour of non-verbal humans; that is, preverbal children, deaf mutes, aphasics, etc. The prevailing view here is to say that such individuals conduct their cognitive business in the language of thought (Mentalese), which is construed literally as a language and one which is not learned, but comes as part of the innate endowment. As may be readily imagined, there are intractable difficulties with this view.<sup>4</sup> Second, it implies a radical discontinuity in the cognitive activity and information processing between verbal humans on the one hand and the rest of the animal kingdom on the other. Evolutionary considerations, on the contrary, would suggest that there are bound to be similarities in how skills are learned, how the world is perceived and represented, how information is stored and retrieved, and so on. Certainly, there will be some differences, but human intelligent behaviour is not likely to have *wholly* different springs and causes than, say, baboon intelligent behaviour, or elephant intelligent behaviour.

In these criticisms, we are, of course, taking the long view of things. There is no question of shelving folk psychology until there is a better theory to replace it with—until, one might say, neuroscience and experimental psychology are far more developed. Moreover, neuroscience needs to make use of prevailing folk psychological concepts in order to climb its way to a position where it can kick the ladder out

from under. Pure bottom-up physiology would surely be pure exasperating folly.

My discussion of folk psychology and its prospects herein presented is highly cursory and justice is not done either to the defenders of folk psychology or to its critics. Still, my aim here is not to provide the full story, but to give the flavour of issues currently in hot debate among philosophers and psychologists, and more importantly, to show that the replacement of folk psychology is a plausible empirical scenario in the reductive future of that theory.

Nevertheless, the idea that humans might ultimately grow up thinking of themselves within the framework of neurophysiological theory does strike some as totally unimaginable, and there are a number of different reasons which fund this view. First, it might be noticed that the futuristic description itself makes use of folk psychology in saying such humans will 'think' of themselves, and so, it might be argued, this implies that retention is inevitable. Not so. The futuristic description is admittedly couched in the vocabulary of the old theory, but it essentially speculates that another theory and vocabulary might be used by future humans. The fact that we can now state those speculations only within the old theory does not guarantee the correctness or the survival of the old theory. It means simply that the old theory is the only one available *now*. Conscious of the poverty of the available descriptive framework, one can envision the possibility of a superior framework, but supplying the detail is not necessary for the vision. In just this way, people in the heyday of vitalism questioned the adequacy of the entire conceptions, without however being able to see just how future humans might understand and explain the 'vitality' of living things.

Second, it may strike some as unimaginable that the categories of belief, desire, thinking, etc., could actually be replaced in our everyday commerce. Notice, though, that in the first instance the replacement is a theoretical reduction, in the sense specified earlier (p. 1042–1043). Whether the reducing theory becomes common coin depends on many non-theoretical factors. The uncritical ease with which ordinary people now talk about what their left or right hemispheres are doing or need, indicates perhaps both the eagerness to apply new theory and the casual facility with which it is picked up. Even supposing the replacement does reach common usage, vestiges of the old theory may linger on, as, for example, have the 'humours' of animal spirit theory, though their original implications have slowly become fainter and feebler. Surely though, there is nothing especially preposterous in the idea that the reducing theory can become widely internalized. At least it is not more preposterous than the shift from a geocentric to a heliocentric to a Newtonian conception of the universe. We have come to accept easily the idea that the earth is a ball tearing around the sun at breathtaking speed, though as Galileo's tormentors rightly noted, it *seems* otherwise. In the end, seeming is not enough,

and because how things seem is so dependent on what we believe, as the beliefs change the seeming will not be far behind.

Third, the possibility that the reduction may be more replacing than retaining may seem unimaginable because introspection tells us that we think and what we think and what desires we have and act upon. Standing back, even briefly, from the overweening power of introspection is the antidote for this argument. We know that when the mind-brain theorizes about the external world—about the nature of matter, of motion, of the sky—that it can err in its theoretical framework and that some observations using the concepts of the framework will in consequence also be in error. Introspection, like all observations, is observation through the lens of theory. Further, there is no reason to suppose that how the brain theorizes about itself (via folk psychology) should be *immune* to error, while how it theorizes about other parts of the world should be so *prone* to error. If the theory we use in seeing ourselves is systematically misconceived, then so also are the observation reports framed in its terms and concepts. To recall an earlier example, observing the heavenly sphere was a straightforward and untroubled affair for the ancients, as seemingly clean of interpretation as our introspection. Within the confines of a different theory, modern man literally sees something different when he looks up. Should humans internalize the theory which reduces folk psychology, then they will literally introspect quite differently conceived goings-on. Moreover, clear evidence is already available that we are often mistaken in our introspective judgements about why a particular decision or choice was made, even though the cases are not pathological, there is no evidence of mendacity and robust confidence accompanies the report.<sup>14</sup> It appears that the subject theorizes about his own motives and desires in explaining his behaviour in much the same way as he theorizes about the behaviour of others. Privileged access to the unvarnished truth about the causes of one's own behaviour is a disabling myth; access is not privileged, nor the truth unvarnished. Additionally, the possibility must be acknowledged that the result of coming to internalize neurophysiological theory may be an enriched and enlightened understanding of oneself and one's behaviour. If we can but provide a rich enough theory, self-understanding may be profoundly enhanced. Accordingly, I do not find the prospect of the replacement of my old ways of understanding what my mind-brain is doing a threatening prospect.

Finally, it will strike some as unimaginable that neuroscience will eventually replace folk psychology because neuroscience, it is said, is concerned with the structural minutiae and the explanation of behaviour will be to advert to higher level capacities and activities. This is an objection voiced more frequently, I suspect, by philosophers and psychologists<sup>17</sup> than by neuroscientists and it is best met by noting that there

is no *methodological* constraint which bars neurophysiological hypotheses about the behaviour of neuronal networks, or which bars the introduction of network properties. Quite the contrary, such hypotheses are devoutly to be sought. Neuroscience can be expected to fashion neurofunctional and neuro-informational concepts which describe activities on a series of levels higher than those which describe the business of individual neurons. Recent work in neuromodelling represents steps in this direction (see Pellionisz & Llinas,<sup>15</sup> Marr,<sup>13</sup> and see my discussion<sup>5</sup>). Whether the details of such hypotheses are sound is not a matter for discussion here; rather, what needs underscoring is that the ascent to network or configurational hypotheses is already underway and this should give pause a-plenty to those who see neuroscience as confined to minutiae-mongering. Perhaps, of course, neuroscience will not succeed in eking out a reductive theory because it might be too hard; the brain might be more complicated than it is smart. But perhaps not, and in any event, it is certainly too soon to announce we have gone about as far as we can go.

#### CONCLUSION

Although it is not possible to predict with any detail the degree of retention which will be exhibited

in the eventual reduction of psychology to neuroscience, what should be stressed here is that there is a spectrum of possibilities and that *a priori* constraints and restrictions about what form the reduction must take will have to content themselves with the hindmost. How the reduction goes is an *empirical* matter and it depends upon empirical results, concerning how in fact the mind-brain works, not on *a priori* intuitions about what is unimaginable. Here, as everywhere else in science, lots of surprises are in store, some of them confounding, dumbfounding and counter-intuitive, and intuitions which will not budge to accommodate tested results will simply be left behind. Current research on the brain—top-down, bottom-up, middle-reaching, omni-directional or whatever—is enormously exciting and I envy profoundly the future scientist who will be much less mysterious to himself than I am to myself.

*Acknowledgements*—I am particularly grateful to P. M. Churchland for help and advice and also to A. W. MacKenzie and N. MacKenzie. This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada, no. 410-81-0182.

#### REFERENCES

1. Causey R. (1976) *The Unity of Science* D. Reidel, Dordrecht.
2. Churchland P. M. (1979) *Scientific Realism and The Plasticity of Mind* Cambridge University Press, Cambridge.
3. Churchland P. M. (1981) Eliminative materialism and the propositional attitudes. *J. Phil.* **78**, 2, 67–90.
4. Churchland P. S. (1978) Fodor on language learning. *Synthese* **38**, 149–159.
5. Churchland P. S. (1980a) A perspective on mind-brain research. *J. Phil.* **77**, 4, 185–207.
6. Churchland P. S. (1980b) Neuroscience and psychology: should the labor be divided? *Behav. Brain Sci.* **3**, 1, 133.
7. Dennett D. C. (1978) *Brainstorms* Bradford/MIT, Cambridge, Mass.
8. Feyerabend P. K. (1958) An attempt at a realist interpretation of experience. *Proc. Aristotlean Soc.* new series.
9. Feyerabend P. K. (1962) Explanation, reduction and empiricism. In *Minnesota Studies in the Philosophy of Science* (eds Fiegl H. & Maxwell G.) Vol. 2. University of Minnesota Press, Minneapolis.
10. Fodor J. A. (1975) *The Language of Thought* Crowell, New York.
11. Hesse M. (1970) Is there an independent observation language? In *The Nature and Function of Scientific Theories* (ed. Colodny R.) University of Pittsburgh Press, Pittsburgh.
- 11a. Hooker, C. A. (1981) Towards a general theory of reduction (I). *Dialogue* **20**, 1, 38–59.
12. Hull D. (1974) *Philosophy of Biological Science* Prentice-Hall, Englewood Cliffs, NJ.
13. Marr D. (1981) Representing visual information. In *Theoretical Approaches In Neurobiology* (eds Reichardt W. & Poggio T.) pp. 151–166. MIT Press, Cambridge Mass.
14. Nisbett R. & Ross L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment* Prentice-Hall, Englewood Cliffs, NJ.
15. Pellionisz A. & Llinas R. (1979) Brain modeling by tensor network theory and computer simulation: distributed processor for predictive coordination. *Neuroscience* **4**, 323–348.
16. Popper K. & Eccles J. (1977) *The Self and its Brain* Springer-Verlag, Berlin.
17. Pylyshyn Z. (1980) Computation and cognition: issues in the foundation of cognitive science *Behav. Brain Sci.* **3**, 1, 111–169.
18. Schaffner K. F. (1974) The unity of science and theory construction in molecular biology. In *Boston Studies in the Philosophy of Science* (eds Seegar R. J. & Cohen R. S.) Vol. 11. D. Reidel, Dordrecht.
19. Sellars W. (1956) Empiricism and the philosophy of mind. In *Minnesota Studies in the Philosophy of Science* (eds Fiegl H. and Scriven M.) Vol. 1. University of Minnesota Press, Minneapolis.
20. Stich S. (1978a) Beliefs and subdoxastic states. *Phil. Sci.* **45**, 4, 499–518.
21. Stich S. (1978b) Autonomous psychology and the belief-desire thesis. *The Monist* **61**, 4, 573–591.
22. Wimsatt W. (1976) Reductive explanation: a functional account. In *PSA 1974* (eds Cohen R. S., Hooker C. A., Michalos A. C. & van Evra J. W.) D. Reidel, Dordrecht.

(Accepted 10 October 1981)