6

Intertheoretic Reduction: A Neuroscientist's Field Guide

Paul M. Churchland and Patricia S. Churchland

Might psychology someday be reduced to (= exhaustively explained by) computational neurobiology? Many still say no. We approach this question through a brief survey of some prominent intertheoretic reductions drawn from our scientific history. A general characterization of reduction is constructed from these, and some important philosophical and methodological lessons are drawn. The five most popular objections to the possibility of a neurobiological reduction of psychology are then addressed and defeated.

"Reductionism" is a term of contention in academic circles. For some, it connotes a right-headed approach to any genuinely scientific field, an approach that seeks intertheoretic unity and real systematicity in the phenomena. It is an approach to be vigorously pursued and defended.

For others, it connotes a wrong-headed approach that is narrow-minded and blind to the richness of the phenomena. It is a bullish instance of "nothing-butery," insensitive to emergent complexity and higher-level organization. It is an approach to be resisted.

One finds this latter reaction most often within the various social sciences, such as anthropology, sociology, and psychology. One finds the former attitude most often within the physical sciences, such as physics, chemistry, and molecular biology. Predictably then, the issue of reductionism is especially turbulent at the point where these two intellectual rivers meet: in the discipline of modern neuroscience.

The question at issue is whether it is reasonable to expect, and to work toward, a reduction of all psychological phenomena to neurobiological and neurocomputational phenomena. A large and still respectable contingent within the academic community remains inclined to say no. Their resistance is principled. Some point to the existence of what philosophers call *qualia*—the various subjective qualitative characters displayed in our sensations: think of pain, the smell of a rose, the sensation of redness, and so forth. These qualia, it is held, are beyond the possibility of any materialist explanation or reduction (Jackson 1982; Nagel 1974). Others point to the semantic content or *intentionality* of our thoughts, and make a similar claim about its irreducibility (Popper and Eccles 1978; Searle 1980a, 1990). Others claim that the most important aspects of human behavior are explicable only in terms of high-level *emergent properties* and their correlative regularities, properties that irreducibly encompass the social level, properties such as loyalty to a moral ideal, perception of a political fact, or the recognition of a personal betrayal (Taylor 1970, 1987). Yet others see a conflict with the important and deeply entrenched idea of *human freedom* (Popper and Eccles 1978). Finally, some materialists raise what is called the problem of *multiple instantiation*. They point to the presumed fact that conscious intelligence could be sustained by physical systems other than the biochemistry peculiar to humans—by a system of transistors, for example—just as a nation's financial economy can be sustained by tokens other than silver coins and paper bills. But no one thinks that macroeconomics can be reduced to the chemistry of metals and paper. So why think that psychology should be reducible to the neurobiology of terrestrial humans? (Fodor 1975).

Our aim in this paper is threefold. First, we try to provide a useful overview of the general nature of intertheoretic reduction, as it appears in the many examples to be found in the history of science. Expanding our horizons here is important, since little is to be learned from simply staring long and hard at the problematic case at issue, namely, the potential reduction of psychological phenomena to neural phenomena. Instead, we need to look at cases where the dust has already settled and where the issues are already clear. Second, we identify the very real virtues that such cases display, and the correlative vices to be avoided. And finally, we attempt to apply these historical lessons to the case here at issue—cognitive neuroscience—and we try to meet the salient objections listed above.

Intertheoretic Reduction: Some Prototypical Cases

A general definition would not be particularly useful at this stage. Since nothing instructs like examples, let us briefly examine some. One of the

Kepler's three planetary laws are:

1. All planets move on ellipses with the sun at one focus.
2. A given planet always sweeps out equal areas in equal times.
3. The square of a planet's period is proportional to the cube of its mean orbital radius.

Newton's three laws of motion are:

1. Inertial motion is constant and rectilinear.
2. Acceleration = force/mass.
3. For any change in momentum, something suffers an equal and opposite change in momentum.

To these laws we must add his gravitation law:

4. $F = Gm_1m_2/R^2$

earliest cases of intertheoretic reduction on a grand scale was the reduction of Kepler's three laws of astronomical motion by the newly minted mechanics of Isaac Newton. Kepler's theory was specific to the motions of the solar planets, but Newton's theory at least purported to be the correct account of bodily motions in general. It was therefore a great triumph when Newton showed that one could deduce all three of Kepler's laws from his own theory, given only the background assumption that the mass of any planet is tiny compared to the great mass of the sun.

Kepler's account thus turned out to be just a special case or a special application of Newton's more encompassing account. And astronomical motions turned out to be just a special instance of the inertial and force-governed motions of massive bodies in general. The divine or supernatural character of the heavens was thereby lost forever. The sublunary and the superlunary realms were thereby united as a single domain in which the same kinds of objects were governed by one and the same set of laws.

Newton's mechanics also provides a second great example of intertheoretic reduction, one that did not emerge until the nineteenth century. If his mechanics successfully comprehends motion at both the astronomical and the human-sized scales, then what, it was asked, about motions at the *microscopic* scale? Might these be accounted for in the same way?

The attempts to construct such an account produced another unification, one with an unexpected bonus concerning the theory of heat. If we assume that any confined body of gas consists of a swarm of submicroscopic corpuscles bouncing around inside the container according to Newton's three laws, then we can deduce a law describing the pressure they will collectively exert on the container's walls by repeatedly bouncing off them. This "kinetic" law has the form

$$PV = 2n/3 \times mv^2/2.$$

This law had the same form as the then already familiar "ideal gas law,"

$$PV = \mu\mathrm{R} \times T.$$

Here P is pressure and V is volume. Although they are notationally different, the expressions $2n/3$ and $\mu\mathrm{R}$ both denote the *amount* of gas present in the container (n denotes the number of molecules in the container; μ denotes the fraction of a mole). The only remaining difference, then, is that the former law has an expression for the *kinetic energy of an average corpuscle* ($mv^2/2$) in the place where the latter has an expression for *temperature* (T). Might the phenomenon we call "temperature" thus *be* mean kinetic energy (KE) at the molecular level? This striking convergence of principle, and many others like it, invited Bernoulli, Joule, Kelvin, and Boltzmann to say yes. Mean molecular kinetic energy turned out to have *all* the causal properties that the classical theory had been ascribing to temperature. In short, temperature turned out to *be* mean molecular KE. Newtonian mechanics had another reductive triumph in hand. Motion at all three scales was subsumed under the same theory, and a familiar phenomenal property, *temperature*, was reconceived in a new and unexpected way.

It is worth emphasizing that this reduction involved identifying a familiar *phenomenal* property of common objects with a highly unfamiliar microphysical property. (By "phenomenal," we mean a property that is reliably discriminated in experience, but where one is unable to articulate, by reference to yet simpler discriminable elements, how one discriminates that property.) Evidently, reduction is not limited to conceptual frameworks hidden away in the theoretical stratosphere. Sometimes the conceptual framework that gets subsumed by a deeper vision turns out to be a familiar piece of our commonsense framework, a piece whose concepts are regularly applied in casual observation on the basis of our native sensory systems. Other examples are close at hand: before

Newton, *sound* had already been identified with compression waves in the atmosphere, and *pitch* with wavelength, as part of the larger reduction of commonsense sound and musical theory to mechanical acoustics. A century and a half after Newton, *light* and its various *colors* were identified with electromagnetic waves and their various wavelengths, within the larger reduction of geometrical optics by electromagnetic theory, as outlined by Maxwell in 1864. *Radiant heat*, another commonsense observable, was similarly reconceived as long-wavelength electromagnetic waves in a later articulation of the same theory. Evidently, the fact that a property or state is at the prime focus of one of our native discriminatory faculties does not mean that it is exempt from possible reconception within the conceptual framework of some deeper explanatory theory (see below).

This fact will loom larger later in the paper. For now, let us explore some further examples of intertheoretic reduction. The twentieth-century reduction of classical (valence) chemistry by atomic and subatomic (quantum) physics is another impressive case of conceptual unification. Here the structure of an atom's successive electron shells, and the character of stable regimes of electron-sharing between atoms, allowed us to reconstruct, in a systematic and thus illuminating way, the electronic structure of the many atomic elements, the classical laws of valence bonding, and the gross structure of the periodic table. As often happens in intertheoretic reductions, the newer theory also allowed us to explain much that the old theory had been unable to explain, such as the specific heat capacities of various substances and the interactions of chemical compounds with light.

This reduction of chemistry to physics is notable for the further reason that it is not yet complete, and probably never will be. For one thing, given the combinatorial possibilities here, the variety of chemical compounds is effectively endless, as are their idiosyncratic chemical, mechanical, optical, and thermal properties. And for another, the calculation of these diverse properties from basic quantum principles is computationally daunting, even when we restrict ourselves to merely approximate results, which for sheerly mathematical reasons we generally must. Accordingly, it is not true that all "chemical" knowledge has been successfully reconstructed in quantum-mechanical terms. Only the basics have, and then only in approximation. But our experience here bids us believe that quantum physics has indeed managed to grasp the underlying elements of chemical reality. We thus expect that any particular *part* of chemistry can be approximately reconstructed in quantum-mechanical terms, when and if the specific need arises. As often it does.

The preceding examples make it evident that intertheoretic reduction is at bottom a relation between two distinct *conceptual frameworks* for describing the phenomena, rather than a relation between two distinct domains of phenomena. The whole point of a reduction, after all, is to show that what we thought to be two domains is actually one domain, though it may have been described in two (or more) different vocabularies.

Perhaps the most famous reduction of all is Einstein's twentieth-century reduction of Newton's three laws of motion by the quite different mechanics of the special theory of relativity (STR). STR subsumed Newton's laws in the following sense. If we make the (false) assumption that all bodies move with velocities much less than the velocity of light, then STR entails a set of laws for the motion of such bodies, a set that is experimentally indistinguishable from Newton's old set. It is thus no mystery that those old Newtonian laws seemed to be true, given the relatively parochial human experience they were asked to account for.

But while those special-case STR laws may be experimentally indistinguishable from Newton's laws, they are logically and semantically quite different from Newton's laws: they ascribe an importantly different family of features to the world. Specifically, in every situation where Newton ascribed an intrinsic property to a body (e.g., mass, or length, or momentum, and so forth), STR ascribes a *relation*, a *two*-place property (e.g., x has a mass-relative-to-an-inertial-frame-F, and so on), because its portrait of the universe and what it contains (a unitary four-dimensional spacetime continuum with 4-D world-lines) is profoundly different from Newton's.

Here we have an example where the special-case resources and deductive consequences of the new and more general theory are not identical, but merely *similar*, to the old and more narrow theory it purports to reduce. That is to say, the special-case reconstruction achieved within the new theory parallels the old theory with sufficient systematicity to explain the old theory's apparent truth, and to demonstrate that the old theory could be displaced by the new without predictive or explanatory loss within the old theory's domain; and yet the new reconstruction is not perfectly isomorphic with the old theory. The old theory turns out not just to be narrow, but to be false in certain important respects. Space and time are not distinct, as Newton assumed, and there simply are no intrinsic properties such as mass and length that are invariant over all inertial frames.

The trend of this example leads us toward cases where the new and more general theory does not sustain the portrait of reality painted by the old theory at all, even as a limiting special case or even in its roughest outlines. An example would be the outright displacement, without reduction, of the old phlogiston theory of combustion by Lavoisier's oxygen theory of combustion. The older theory held that the combustion of any body involved the *loss* of a spirit-like substance, phlogiston, whose precombustion function it was to provide a noble wood-like or metal-like character to the baser ash or calx that is left behind after the process of combustion is complete. It was the "ghost" that gave metal its form. With the acceptance of Lavoisier's contrary claim that a sheerly material substance, oxygen, was being somehow *absorbed* during combustion, phlogiston was simply eliminated from our overall account of the world.

Other examples of theoretical entities that have been eliminated from serious science include caloric fluid, the rotating crystal spheres of Ptolemaic astronomy, the four humors of medieval medicine, the vital spirit of premodern biology, and the luminiferous aether of pre-Einsteinian mechanics. In all of these cases, the newer theory did not have the resources adequate to reconstruct the furniture of the older theory or the laws that supposedly governed its behavior; but the newer theory was so clearly superior to the old as to displace it regardless.

At one end of the spectrum then, we have pairs of theories where the old is smoothly reduced by the new, and the ontology of the old theory (that is, the set of things and properties that it postulates) survives, although redescribed, perhaps, in a new and more penetrating vocabulary. Here we typically find claims of cross-theoretic identity, such as "Heat is identical with mean molecular kinetic energy" and "Light is identical with electromagnetic waves." In the middle of the spectrum, we find pairs of theories where the old ontology is only poorly mirrored within the vision of the new, and it "survives" only in a significantly modified form. Finally, at the other end of the spectrum we find pairs where the older theory, and its old ontology with it, is eliminated entirely in favor of the more useful ontology and the more successful laws of the new.

Before closing this quick survey, it is instructive to note some cases where the older theory is neither subsumed under nor eliminated by the aspirant and allegedly more general theory. Rather, it successfully resists the takeover attempt, and proves not to be just a special case of the general theory at issue. A clear example is Maxwell's electromagnetic (hereinafter EM) theory. From 1864 to 1905, it was widely expected that EM theory would surely find a definitive reduction in terms of the mechanical properties of an all-pervading aether, the elastic medium in which EM waves were supposedly propagated. Though never satisfactorily completed, some significant attempts at reconstructing EM phenomena in mechanical terms had already been launched. Unexpectedly, the existence of such an absolute medium of luminous propagation turned out to be flatly inconsistent with the character of space and time as described in Einstein's 1905 special theory of relativity. EM theory thus emerged as a fundamental theory in its own right, and not just as a special case of mechanics. The attempt at subsumption was a failure.

A second example concerns the theory of stellar behavior accumulated by classical astronomy in the late nineteenth century. It was widely believed that the pattern of radiative behavior displayed by a star would be adequately explained in mechanical or chemical terms. It became increasingly plain, however, that the possible sources of chemical and mechanical energy available to any star would sustain their enormous outpourings of thermal and luminous energy for no more than a few tens of millions of years. This limited time scale was at odds with the emerging geological evidence of a history numbered in the *billions* of years. Geology notwithstanding, Lord Kelvin himself was prepared to bite the bullet and declare the stars to be no more than a few tens of millions of years old. The conflict was finally resolved when the enormous energies in the atomic nucleus were discovered. Stellar astronomy was eventually reduced all right, and very beautifully, but by quantum physics rather than by mere chemistry or mechanics. Another reductive attempt had failed, though it was followed by one that succeeded.

The Lessons for Neuroscience

Having seen these examples and the spectrum of cases they define, what lessons should a neuroscientist draw? One lesson is that intertheoretic reduction is a normal and fairly commonplace event in the history of science. Another lesson is that genuine reduction, when you can get it, is clearly a *good* thing. It is a good thing for many reasons, reasons made more powerful by their conjunction. First, by being displayed as a special case of the (presumably true) new theory, the old theory is thereby *vindicated*, at least in its general outlines, or at least in some suitably restricted domain. Second, the old theory is typically *corrected* in some of its important details, since the reconstructed image is seldom a perfect mirror image of the old theory, and the differences reflect improvements in our knowledge. Third, the reduction provides us with a much *deeper insight* into, and thus a *more effective control* over, the phenomena within the old theory's domain. Fourth, the reduction provides us with a *simpler* overall account of nature, since apparently diverse phenomena are brought under a single explanatory umbrella. And fifth, the new and more general theory immediately *inherits all the evidence* that had accumulated in favor of the older theory it reduces, because it explains all of the same data.

It is of course a bad thing to try to force a well-functioning old theory into a Procrustean bed, to try to effect a reduction where the aspirant reducing theory lacks the resources to do reconstructive justice to the target old theory. But whether or not the resources are adequate is seldom clear beforehand, despite people's intuitive convictions. And even if a reduction is impossible, this may reflect the old theory's radical falsity instead of its fundamental accuracy. The new theory may simply eliminate the old, rather than smoothly reduce it. Perhaps folk notions such as "beliefs" and the "will," for example, will be eliminated in favor of some quite different story of information storage and behavior initiation.

The fact is, in the neuroscience-psychology case there are conflicting indications. On the one side, we should note that the presumption in favor of an eventual reduction (or elimination) is far stronger than it was in the historical cases just examined. For unlike the earlier cases of light, or heat, or heavenly motions, in general terms we already know how psychological phenomena arise: they arise from the evolutionary and ontogenetic articulation of matter, more specifically, from the articulation of biological organization. We therefore *expect* to understand the former in terms of the latter. The former is produced by the relevant articulation of the latter.

But there are counterindications as well, and this returns us at last to the five objections with which we opened this paper. From the historical perspective outlined above, can we say anything useful about those objections to reduction? Let us take them in sequence.

The first concerns the possibility of explaining the character of our subjective sensory qualia. The negative arguments here all exploit the very same theme, viz., our inability to imagine how any possible story about the objective nuts and bolts of neurons could ever explain the inarticulable subjective phenomena at issue. Plainly this objection places a great deal of weight on what we can and cannot imagine, as a measure of what is and is not possible. It places more, clearly, than the test should bear. For who would have imagined, before James Clerk Maxwell, that the theory of charged pith balls and wobbling compass needles could prove adequate to explain all the phenomena of light? Who would have thought, before Descartes, Bernoulli, and Joule, that the mechanics of billiard balls would prove adequate to explain the prima facie very different phenomenon of heat? Who would have found it remotely plausible that the pitch of a sound is a frequency, in advance of a general appreciation that sound itself consists in a train of compression waves in the atmosphere?

We must remember that a successful intertheoretic reduction is typically a complex affair, as it involves the systematic reconstruction of all or most of the old conception within the resources of the new conception. And not only is it complex, often the reconstruction is highly surprising. It is not something that we can reasonably expect anyone's imagination to think up or comprehend on rhetorical demand, as in the question, How could A's *possibly* be nothing but B's?

Besides, an imagination informed by recent theories of sensory coding need not be so stumped as the rhetorical question expects. The idea that taste sensations are coded as a four-dimensional vector of spiking frequencies (corresponding to the four types of receptor on the tongue) yields a representation of the space of humanly possible tastes which unites the familiar tastes according to their various similarities, differences, and other relations such as betweenness (Bartoshuk 1978). Land's retinex theory of color vision (Land 1977) suggests a similar arrangement for our color sensations, with similar virtues. Such a theory also predicts the principal forms of color blindness, as when one's three-dimensional color space is reduced to two dimensions by the loss of one of its normal dimensions of representation.

Here we are already reconstructing some of the features of the target phenomena in terms of the new theory. We need only to carry such a reconstruction through, as in the historical precedents of the objective phenomenal properties noted earlier (heat, light, pitch). Some things may indeed be inarticulably phenomenal in character, because they are the target of one of our basic discriminatory modalities. But that in no way makes them immune to an illuminating intertheoretic reduction. History already teaches us the contrary.

The second objection concerned the meaning, or semantic content, or intentionality of our thoughts and other mental states. The antireductionist arguments in this area are very similar to those found in the case of qualia. They appeal to our inability to imagine how meaning could be just a matter of how signals interact or how inert symbols are processed. (Searle 1980a, 1990; for a rebuttal, see P. M. Churchland and P. S. Churchland 1990b. Searle, strictly speaking, objects only to a purely computational reduction, but that is an important option for neuroscience so we shall include him with the other antireductionists.) Such appeals, as before, are really arguments from ignorance. They have the form, I can't *imagine* how a neurocomputational account of meaningful representations could possibly work; therefore, it can't possibly work. To counter such appeals in the short term, we need only point this failing out.

To counter them in the long term requires more. It requires that we actually produce an account of how the brain represents the external world and the regularities it displays. But that is precisely what current theories of neural network function address. Real-time information about the world is coded in high-dimensional activation vectors, and general information about the world is coded in the background configuration of the network's synaptic weights. Activation vectors are processed by the weight configurations through which they pass, and learning consists in the adjustment of one's global weight configuration. These accounts already provide the resources to explain a variety of things, such as the recognition of complex objects despite partial or degraded sensory inputs, the swift retrieval of relevant information from a vast content-addressable memory, the appreciation of diffuse and inarticulable similarities, and the administration of complex sensorimotor coordination (P. M. Churchland 1989a). We are still too ignorant to insist that hypotheses of this sort will prove adequate to explain all of the representational capacities of mind. But neither can we insist that they are doomed to prove inadequate. It is an empirical question, and the jury is still out.

The third objection complains that what constitutes a human consciousness is not just the intrinsic character of the creature itself, but also the rich matrix of relations it bears to the other humans, practices, and institutions of its embedding culture. A reductionistic account of human consciousness and behavior, insofar as it is limited to the microscopic activities in an individual's brain, cannot hope to capture more than a small part of what is explanatorily important.

The proper response to this objection is to embrace it. Human behavior is indeed a function of the factors cited. And the character of any individual human consciousness will be profoundly shaped by the culture in which it develops. What this means is that any adequate neurocomputational account of human consciousness must take into account the manner in which a brain comes to represent not just the gross features of the physical world but also the character of the other cognitive creatures with which it interacts, and the details of the social, moral, and political world in which they all live. The brains of social animals, after all, learn to be interactive elements in a community of brains, much to their cognitive advantage. We need to know how they do it.

This is a major challenge, one that neuroscientists have not yet addressed with any seriousness, nor even much acknowledged. This is not surprising. Accounting for a creature's knowledge of the spatial location of a fly is difficult enough. Accounting for its knowledge of a loved one's embarrassment, a politician's character, or a bargaining opponent's hidden agenda, represents a much higher level of difficulty. And yet we already know that artificial neural networks, trained by examples, can come to recognize and respond to the most astonishingly subtle patterns and similarities in nature. If physical patterns, why not social patterns? We confront no problem in principle here. Only a major challenge.

It may indeed be unrealistic to expect an exhaustive global account of the neural and behavioral trajectory of a specific person over any period of time. The complexity of the neural systems we are dealing with may forever preclude anything more than useful approximations to the desired ideal account. The case of chemistry and its relation to quantum-physics comes to mind. There also, the mathematics of complex dynamical systems imposes limits on how easily and accurately we can reconstruct the chemical facts from the physical principles. This means that our reduction will never be truly complete, but we rightly remain confident that chemical phenomena are nothing but the macrolevel reflection of the underlying quantum-physical phenomena even so. As with chemical phenomena, so with psychological phenomena.

This brings us to the fourth objection, concerning the threat that a reduction would pose to human freedom. Here we shall be brief. Whether and in what sense there is any human freedom, beyond the relative autonomy that attaches to any complex dynamical system that is partially isolated from the world, is an entirely empirical question. Accordingly, rather than struggle to show that a completed neuroscience will be consistent with this, that, or the other preconceived notion of human freedom, we recommend that we let scientific investigation *teach us* in what ways and to what degrees human creatures are "free." No doubt this will entail modifications for some people's current conceptions of human freedom, and the complete elimination of some others. But that is preferable to making our current confusions into a standard that future theories must struggle to be consistent with.

The fifth and final objection claims an irreducibly abstract status for psychology, on grounds that a variety of quite different physical systems could realize equally well the abstract organization that constitutes a cognitive economy. How can we reduce psychological phenomena to neurobiology, if other physical substrates might serve just as well?

The premise of this objection will likely be conceded by all of us. But the conclusion against reduction does not follow. We can see this clearly by examining a case from our own scientific history. Temperature, we claimed earlier, is identical with mean molecular kinetic energy. But strictly speaking, this is true only for a gas, where the molecules are free to move in a ballistic fashion. In a solid, where the particles oscillate back and forth, their energy is constantly switching between a kinetic and a potential mode. In a high-temperature plasma, there are no molecules at all to consider, since everything has been ripped into subatomic parts. Here temperature is a complex mix of various energies. And in a vacuum, where there is no mass at all, temperature consists in the wavelength distribution—the "black-body curve"—of the EM waves passing through it.

What these examples show us is that reductions can be "domain specific": in a gas, temperature is one thing; in a solid, temperature is another thing; in a plasma, it is a third; in a vacuum, a fourth; and so on. (They all count as temperatures, since they interact, and they all obey the same laws of equilibrium and disequilibrium.) None of this moves us to say that classical thermodynamics is an autonomous, irreducible science, forever safe from the ambitions of the underlying microphysical story. On the contrary, it just teaches us that there is more than one way in which energy can be manifested at the microphysical level.

Similarly, visual experience may be one thing in a mammal, and a slightly different thing in an octopus, and a substantially different thing in some possible metal-and-semiconductor android. But they will all count as visual experiences because they share some set of abstract features at a higher level of description. That neurobiology should prove capable of explaining all psychological phenomena in humans is not threatened by the possibility that some *other* theory, say semiconductor electronics, should serve to explain psychological phenomena in *robots*. The two reductions would not conflict. They would complement each other.

We have elsewhere provided more comprehensive accounts of how recent work in neuroscience illuminates issues in psychology and cognitive theory (P. S. Churchland 1986; P. M. Churchland 1989a). Let us conclude this paper with two cautionary remarks. First, while we have here been very upbeat about the possibility of reducing psychology to neuroscience, producing such a reduction will surely be a long and difficult business. We have here been concerned only to rebut the counsel of impossibility, and to locate the reductive aspirations of neuroscience in a proper historical context.

Second, it should not be assumed that the science of psychology will somehow disappear in the process, nor that its role will be limited to that of a passive target of neural explanation. On the contrary, chemistry has not disappeared despite the quantum-mechanical explication of its basics; nor has the science of biology disappeared, despite the chemical explication of its basics. And each of these higher-level sciences has helped to shape profoundly the development and articulation of its underlying science. It will surely be the same with psychology and neuroscience. At this level of complexity, intertheoretic reduction does not appear as the sudden takeover of one discipline by another; it more closely resembles a long and slowly maturing marriage.