THE NEW ATLANTIS
A JOURNAL OF TECHNOLOGY & SOCIETY

The Limits of Neuro-Talk

Matthew B. Crawford

> If he be a man engaged in any important inquiry, he must have a method, and he will be under a strong and constant temptation to make a metaphysics out of his method, that is, to suppose the universe ultimately of such a sort that his method must be appropriate and successful.
>
> –E. A. Burtt, *The Metaphysical Foundations of Modern Science* (1925)

In this nascent age of "neurolaw," "neuromarketing," "neuropolicy," "neuroethics," "neurophilosophy," "neuroeconomics," and even "neurotheology," it becomes necessary to disentangle the science from the scientism. There is a host of cultural entrepreneurs currently grasping at various forms of authority through appropriations of neuroscience, presented to us in the corresponding dialects of neuro-talk. Such talk is often accompanied by a picture of a brain scan, that fast-acting solvent of critical faculties.

Elsewhere in this issue, O. Carter Snead offers a critique of the use of brain scans in the courtroom in which he alludes to, but ultimately brackets, questions about the scientific rigor of such use. For the sake of argument, he proceeds on the assumption that neuroimaging is competent to do what it is often claimed to do, namely, provide a picture of human cognition.

But there are some basic conceptual problems hovering about the interpretation of brain scans as pictures of mentation. In parsing these problems, it becomes apparent that the current "neuro" enthusiasm should be understood in the larger context of scientism, a pervasive cultural tendency with its own logic. A prominent feature of this logic is the overextension of some mode of scientific explanation, or model, to domains in which it has little predictive or explanatory power. Such a lack of intrinsic fit is often no barrier to the model nonetheless achieving great authority in those domains, through a kind of histrionics. As Alasdair MacIntyre has shown in another context (that of social science), all that is required is a certain kind of performance by those who foist the model upon us, a dramatic *imitation* of explanatory competence that wows us and cows us with its self-confidence. At such junctures, the heckler performs an important public service.

Matthew B. Crawford *is a fellow at the Institute for Advanced Studies in Culture at the University of Virginia and a contributing editor of* The New Atlantis.

Taxonomies of Mind

As applied to medical diagnosis (for example, in diagnosing a brain tumor), a brain scan is similar in principle to a mammogram: it is a way of seeing inside the body. Its success at doing so is straightforward and indubitable. However, the use of neuroimaging in psychology is a fundamentally different kind of enterprise: it is a research method the validity of which depends on a premise. That premise is that mental processes can be analyzed into separate and distinct faculties, components, or modules, and further that these modules are instantiated, or realized, in localized brain regions. Jerry Fodor, the Rutgers University philosopher of mind, begins his classic 1983 monograph *The Modularity of Mind* thus:

> Faculty psychology is getting to be respectable again after centuries of hanging around with phrenologists and other dubious types. By faculty psychology I mean, roughly, the view that many fundamentally different kinds of psychological mechanisms must be postulated in order to explain the facts of mental life. Faculty psychology takes seriously the apparent heterogeneity of the mental and is impressed by such *prima facie* differences as between, say, sensation and perception, volition and cognition, learning and remembering, or language and thought.

With its due regard for the heterogeneity of our mental experience, this modularity thesis is indeed attractive as a working hypothesis. The difficulty lies in arriving at a specific taxonomy of the mental. The list of faculties Fodor gives in the paragraph above could be replaced with an indefinite number of competing taxonomies—and indeed, Fodor gives a taxonomy of taxonomies. The discipline of psychology exhibits a lack of agreement on the most basic elements of the mental.

The problem of classifying the mental is one that infects the neuroimaging enterprise at its very roots. Some observers argue this problem is fatal to the interpretation of brain images as representing well-defined cognitive processes. One such critic is William Uttal, a psychologist at Arizona State University. In his 2001 book *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*, Uttal shows that there has been no convergence of mental taxonomies over time, as one might expect in a mature science. "Rather," he writes, "a more or less expedient and highly transitory system of definitions has been developed in each generation as new phenomena are observed or hypothetical entities created."

Uttal suggests that the perennial need to divide psychology textbooks into topic chapters—"pattern recognition," "focal attention," "visual

memory," "speech perception," and the like—has repeatedly induced an unwitting reification of such terms, whereby they come to be understood as separable, independent modules of mental function. The ad hoc origin of such mental modules subsides from the collective memory of investigators, who then set out to search for their specific loci in the brain.

Ideally, the phenomenological work of arriving at a taxonomy of the mental would be accomplished *prior* to the effort to tie mental functions to brain regions, because without such a taxonomy in hand, there is a real risk that arbitrary features of the imaging technology will yield artifacts that may then, like textbook categories, get reified as mental modules. Such artifacts are just the tip of a metaphysical iceberg of the sort Burtt warned of in the epigraph above.

Moreover, an even more fundamental problem haunts the modular theory of mind assumed in cognitive neuroscience: it may be that neither mental functions, nor the physical systems that realize them, are decomposable into independent modules. Uttal argues that rather than distinct *entities*, the various features of cognition are likely to be *properties* of a more general mental activity that is distributed throughout the brain. For example, is it possible neatly to distinguish perception from attention? Uttal asks of attention,

> Is it a "stuff" that can be divided, allocated, and focused and that is available only in limited amounts, and thus can be localized in a particular part of the brain? Or, to the contrary, is it an attribute or characteristic of perception…inseparable as the diameter or whiteness of a golf ball is from the physical ball itself?…It seems plausible that many of the psychological components or modules we seek to locate in a particular region of the brain should be thought of as properties of a unified mental "object" rather than as analyzable and isolatable entities.

This argument is, perhaps, a bit tendentious—who in the neuroimaging literature suggests attention is a "stuff"? Rather, attention is thought to be a *function* realized in some brain region. But this correction does not vacate the force of Uttal's criticism, because functions, like properties, are distributed (they require a whole system or mechanism to be realized), and one must pause to ask: what are the boundaries of the pertinent system? A danger inherent in the localization thesis may be illuminated by analogy to an internal combustion engine. In describing an engine, one might be tempted to say, "the opening of the intake valve is caused by the movement of the rocker arm." Except that the rocker is, in turn, set in motion by

the camshaft, the camshaft by the crankshaft, the crank by a connecting rod, the rod by the piston. But of course, the piston won't move unless the intake valve opens to let the air-fuel mixture in. This logic is finally circular because, really, it is the *entire* mechanism that "causes" the opening of the intake valve; any less holistic view truncates the causal picture and issues in statements that are, at best, partially true. Given that the human brain is more complexly interconnected than a motor by untold orders of magnitude, it is a dubious undertaking to say that any localized organic structure is the sufficient cause and exclusive locus of something like "reason" or "emotion."

Such dichotomous mental categories are regularly employed by social scientists who have taken up neuro-talk, and in the popular press: the amygdala is said to be the seat of emotion, the prefrontal cortex of reason. Yet when I get angry, for example, I generally do so for a *reason*; typically I judge myself or another wronged. To cleanly separate emotion from reason-giving makes a hash of human experience, and seems to be attractive mainly as a way of rendering the mind methodologically tractable, even if at the cost of realism.

Such naïve psychological modularity can get installed in institutional practices that have real coercive power, like lie-detection. There are now at least two companies, No Lie MRI, based in San Diego, and Cephos, based in Boston, that are actively commercializing the application of neuroimaging to lie detection. Margaret Talbot, writing in *The New Yorker* in 2007, described the neuroimaging studies of deception conducted by Daniel Langleben, a psychiatrist at the University of Pennsylvania, that undergird the claims of No Lie MRI: "These allegedly seminal studies look exclusively at…people who were instructed to lie about trivial matters in which they had little stake. An incentive of twenty dollars can hardly compare with, say, your freedom, reputation, children, or marriage—any or all of which might be at risk in an actual lie-detection scenario."

This is to treat lying as a "cognitive" process in the narrowest sense, as opposed to a mental act with inherent ethical content and pragmatic consequences. Here cognitive science reveals its roots in "the linguistic turn" in philosophy that began with the rise of logical positivism a century ago. The logical positivists were preoccupied with consistency of sentences, and conceived reason to be syntactical or rule-like. It is what computers do. Such a view takes no account of what Henri Bergson called "the tension of consciousness," that feature of an embodied being who has *interests* and finds himself situated in a *world*. Talbot nicely elaborates the richness

of the phenomena we gather under the term "lie," and the problem it poses for any narrowly cognitive scheme of lie detection:

> small, polite lies; big, brazen, self-aggrandizing lies; lies to protect or enchant our children; lies that we don't really acknowledge to ourselves as lies; complicated alibis that we spend days rehearsing. Certainly, it's hard to imagine that all these lies will bear the identical neural signature. In their degrees of sophistication and detail, their moral weight, their emotional valence, lies are as varied as the people who tell them. As Montaigne wrote, "The reverse side of the truth has a hundred thousand shapes and no defined limits."

Trying to identify a universal, merely formal element of real-life lying and disentangle it from emotional capacities, moral dispositions, and worldly situations, on the supposition that the function "lie" has its own distinct ontology, may make as much sense as trying to separate the whiteness of a golf ball from the ball, to use Uttal's analogy. The thesis of mental modularity seems to be attractive mainly because it is convenient for talking about thinking and, as we shall see, for designing experiments. But notice that it can also undergird technologies such as brain-scan lie detection that may come to have real consequences in the world—affecting "your freedom, reputation, children, or marriage," as Talbot reminds us. Just because such technologies aren't adequate to our mental reality doesn't mean they won't be deployed; the checkered history of past lie detection technologies shows this. It is significant that No Lie MRI solicits inquiries from corporate customers on its website. Even if the technology doesn't pass the bar of public trust for use by civil authorities, it may satisfy corporate managers looking for new ways to intimidate employees.

Those who would use science to solve real human problems often must first translate those human problems into narrowly technical problems, framed in terms of some theoretically tractable model and a corresponding method. Such tractability offers a collateral benefit: the intellectual pleasure that comes with constructing and tinkering with the model. But there is then an almost irresistible temptation to, as E. A. Burtt said, turn one's method into a metaphysics—that is, to suppose the world such that one's method is appropriate to it. When this procedure is applied to human beings, the inevitable result is that the human is defined downward. Thus, for example, thinking becomes "information processing." We are confronted with the striking reversal wherein cognitive science looks to the computer to understand what human thinking is.

Deep Problems of Instrumentation

If the critique of mental modularity is valid, how can one account for the fact that brain scans do, in fact, reveal well-defined areas that "light up" in response to various cognitive tasks? In the case of functional (as opposed to structural) neuroimaging, what you are seeing when you look at a brain scan is the result of a subtraction. Functional magnetic resonance imaging (fMRI), for example, produces a map of the rate of oxygen use in different parts of the brain, which stands as a measure of metabolic activity. Or rather, it depicts the *differential* rate of oxygen use: one first takes a baseline measurement in the control condition, then a second measurement while the subject is performing some cognitive task. The baseline measurement is then subtracted from the on-task measurement. The reasoning, seemingly plausible, is that whatever shows up in the subtraction represents the metabolic activity associated solely with the cognitive task in question.

One immediately obvious (but usually unremarked) problem is that this method eliminates from the picture the more massive fact, which is that the entire brain is active in both conditions. A false impression of neat functional localization is given by the presentation of differential brain scans which subtract out all the distributed functions. This subtractive method is ideally suited to the imaging technology, and deeply consistent with the modular theory of mind. But is this modular theory of mind perhaps attractive in part *because* it lends itself to the subtractive method?

In the late 1990s and early 2000s, some of the more critical cognitive neuroscientists complained that researchers were simply sticking people in the magnet to see what "lights up," with no real theory in hand, and such studies would get published in prominent scientific journals. These critiques had some effect, and the discipline has mostly moved beyond looking for "the spot for X." Indeed, cognitive neuroscientists deserve credit for the methodological finesse they have developed in response to the complexity of mind.

In a 1999 article in *Behavioral and Brain Sciences*, Cambridge neuroscientist Friedemann Pulvermuller offered a thorough account of the difficulties that arise in the effort to localize linguistic functions. The problem with the subtractive method, he wrote, is that "in many experiments there are [not one, but] several differences between critical and control conditions," on such dimensions as perception (a word is seen or not on a screen), attention, classification (the word may be a noun or verb or meaningless pseudo-word, for example), motor response (the subject

may be required to hit a button as part of his or her response), search processes (the subject may need to recall the word), and semantic inferences. Given that "an area is found to 'light up'…it is not clear which of the many different cognitive processes relates to the difference in brain activity." Similarly, Michele T. Diaz and Gregory McCarthy write in the November 2007 issue of the *Journal of Cognitive Neuroscience* that "the coactivation of cognitive processes unrelated to word processing *per se*… likely influence[s] the pattern of activation obtained in even the simplest word processing tasks."

University of Chicago "social neuroscientist" John T. Cacioppo and colleagues offered a philosophically sophisticated treatment of these methodological hazards of neuroimaging in a 2003 article in the *Journal of Personality and Social Psychology*. They describe as a "category error" the

> intuitively appealing notion that the organization of cognitive phenomena maps in a 1:1 fashion into the organization of the underlying neural substrates. Memories, emotions, and beliefs, for instance, were each once thought to be localized in a single site in the brain. Current evidence, however, now suggests that most complex psychological or behavioral concepts do not map into a single center in the brain. What appears at one point in time to be a singular construct (e.g., memory), when examined in conjunction with evidence from the brain (e.g., lesions) reveals a more complex and interesting organization at both levels (e.g., declarative vs. procedural memory processes). Even if there is localization, it will likely be elusive until there are coherent links between psychological–behavioral constructs and neural operations.

As these articles indicate, the problem of distributed, mutually intertwined mental functions is very much at issue in the trenches of neuroscience, however much grand theorists may find it expedient to ignore such difficulties and insist, as Steven Pinker does in *How the Mind Works*, that "the mind is organized into modules or mental organs, each with a specialized design that makes it an expert in one arena of interaction with the world." Simplifications like this are not culturally innocent, as they provide the indispensable pretext for the grab at authority by entrepreneurs such as No Lie MRI, which in turn may come to justify the exercise of coercive power by civil authorities.

Perhaps the most fundamental limitation of functional imaging, vis-à-vis the claim that it allows us to "peer inside the mind," is that there is a basic disconnect of time scale. Brain scans are emphatically *not* images of cognition in process, as the neural activity of interest occurs on a time

scale orders of magnitude faster than hemodynamic response (the proxy for neural activity measured by fMRI). Uttal writes,

> This raises, once again, a profoundly disconcerting problem for the users of imaging procedures: the cumulative measure of brain metabolism is neither theoretically nor empirically linked to the momentary details of the neural network at the micro level—the essential level of information processing that is really the psychoneural equivalent of mentation. From this point of view, the "signs" of brain activity obtained from the scanning system are no more "codes" of what is going on than any other physiological correlate, such as the electrodermal response or an electromyogram.

I take Uttal to mean that a brain image provided by fMRI may serve as a *sign* of mentation, but because of the time-scale difference, it does not preserve the machine states that (on the computational theory of mind) *encode* mentation. With such signs, we do not have a picture *of* a mechanism. We have a sign *that there is* a mechanism. But the discovery that there is a mechanism is no discovery at all, unless one was previously a dualist.

Respect the Machine

But suppose imaging technology were one day to achieve both the spatial and temporal resolution to give us a precise picture, down to the neuronal level, of the physical correlates of mentation as it occurs. What then? To fully understand even a simple mechanism can be a surprisingly elusive undertaking, as is known by anyone who has—to use another engineering example—tried to set up a train of beveled gears that transmit a rotary motion through a right angle (as in the Ducati motorcycle engines of a few decades ago). In such a gear train there are forces acting in directions that do not correspond to any of the observed motions. There are thrust and side forces that are *intellectually* manageable (they can be expressed mathematically) but *practically* far from trivial. An experienced engine builder may require an entire day playing with shims and tolerances to get it right.

Though beveled gears are only barely more complex than the simple machines of Archimedes (the lever, the pulley, etc.), the human race had to await the genius of Leonardo to receive them, like some revelation. At a much higher level of sophistication, mechanisms can be intractably complex things—famously, the most subtle applications of science and engineering have as yet been unable to fully reproduce the prized characteristics of Stradivari's violins, for example. The human brain, everyone agrees, presents complexity that is simply colossal by comparison—by

one estimate, the number of possible neuronal pathways is larger than the number of particles in the universe.

But for a certain kind of intellectual, the mere act of *positing* that some mystery has a mechanical basis gives satisfaction. A heady feeling of mastery rushes in prematurely with the idea that *in principle* nothing lies beyond our powers of comprehension. But to be knowable in principle is quite different from being known in fact. Hands-on mechanical experience frequently induces an experience of perplexity in formally trained engineers. We may be emboldened to speculate, in a sociological mode, whether a lack of such mechanical experience "enables" a certain intellectual comportment which doesn't give the machine its due, and isn't sufficiently impressed with this difference between the knowable and the known.

A species of metaphysical enthusiast can often be seen skipping gaily across this gap between the knowable and the known, acting in the capacity of publicist for some research program which, on cooler analysis, is seen to be in its infancy. One such is Paul Churchland of the University of California, San Diego, who writes in his 1995 book *The Engine of Reason, the Seat of the Soul* that "we are now in a position to explain how our vivid sensory experience arises in the sensory cortex of our brains… [and is] embodied in a vast chorus of neural activity.… [W]e can now understand how the infant brain slowly develops a framework of concepts…and how the matured brain deploys that framework almost instantaneously: to recognize similarities, to grasp analogies, and to anticipate both the immediate and the distant future." As Jerry Fodor succinctly put it in a review of Churchland's book, "none of this is true"; it is "the idiom of grant proposals." The critical element of Churchland's hype lies in the distinction between knowing *that* "our vivid sensory experience arises in the sensory cortex" and explaining *how* it does so, which latter, he claims, is now accomplished. We surely know *that* "the infant brain slowly develops a framework of concepts" and "the matured brain deploys that framework almost instantaneously," but *how*? Not even to a first glimmer, as Fodor says.

Of Dogs and Protons

The conceit that brain scans present an image of human cognition laid out before us for full inspection holds obvious attraction. This positive appeal is backed up by a horror at what is thought to be the only alternative to a thoroughly reductive materialism: some form of spiritualism or, more broadly, something "anti-scientific."

But one must make a distinction between ontological reduction and explanatory reduction. This distinction is a commonplace in the philosophy of science, but it is routinely ignored in the hype surrounding cognitive neuroscience. The error goes like this: from the fact that some phenomenon is composed of and dependent upon more fundamental parts, it is thought to follow that any explanation of the higher-level phenomenon can be replaced by, or translated without residue into, an explanation at the lower level of its parts. Once this reduction is (putatively) accomplished, the ontological status of the higher-level phenomenon is demoted to that of *mere* phenomenon: appearance versus reality. Our gaze is shifted away from the thing we initially wanted to understand, to some underlying substrate. This procedure is thought to be enjoined by the conviction we all share with the natural scientist: there is only one universe, and it is made up of physical particles.

Yet the natural scientist knows just as surely that our best account of that universe is, in many cases, not forthcoming from physics. We turn instead to chemistry or biology. The need for such "special" sciences that take higher-level structures as given does not compromise the bedrock ontological supposition that there is a single universe, made up of physical particles. One can have one's materialism while admitting the autonomy of higher-level disciplines. There is much confusion on this point, and it seems to be bolstered by a fear that to be less than completely reductive in one's explanatory posture somehow commits one to "spiritualism."

The explanatory independence of biology, its irreducibility to physics, is consistent with biological entities being composed of and dependent upon physical entities. The biologist believes that the dog is made up of nothing but protons, neutrons, and electrons, but he does not try to give an account of the dog at that level. Is this merely due to the limitations of our current state of knowledge? Would it be possible *in principle* to construct a comprehensive understanding of the dog starting from particle physics? The consensus view appears to be that it is not possible even in principle, due to considerations of complexity and non-linearity, or thermodynamic irreversibility (take your pick). Even within physics, lower-level accounts sometimes presuppose structure that is identifiable only at a higher level, or depend upon boundary-conditions that cannot be generated from within the lower-level account. Even something as simple as a volume of gas displays "emergent properties" (here, an irreversible tendency toward equilibrium) that cannot be derived from the collisions between individual gas molecules (which are symmetric with respect to time).

It seems to be not scientists, but rather publicists of science, who are haunted by a sense of metaphysical hazard when confronted with

phenomena that can't be fully understood reductively. But how sincere is this horror? Is it rather a pose struck by such publicists, a histrionic display intended to cow others into submission? "You're not a *dualist*, are you?" For their part, many humanists aren't sufficiently acquainted with the principles of scientific explanation to be able to see that this kind of bullying is fraudulent in its claim to speak for science, and end up feeling resentful towards science. This is bad for humanists, and bad for science.

The Histrionics of the Neuro-Metaphysician

A paper recently published in the *Journal of Cognitive Neuroscience*, of all places, shines a light on the magical, totemic effect of brain scans on those viewing them. The authors of "The Seductive Allure of Neuroscience Explanations," a team of Yale scholars, offered their subjects various explanations for certain psychological phenomena that are familiar to everyday experience. Some of these explanations were contrived to be pointedly *bad* explanations. Their subjects consisted of three groups: neuroscientists, neuroscience students, and lay adults. The study found that all three groups did well at identifying the bad explanations as bad, except when those explanations were preceded with the words, "Brain scans indicate." Then the students and lay adults tended to accept the bad explanation. A complementary set of experiments by David P. McCabe and Alan D. Castel, currently in press in the journal *Cognition*, found that "readers infer more scientific value for articles including brain images than those that do not, regardless of whether the article included reasoning errors or not."

These findings suggest that we are culturally predisposed to surrender our own judgment in the face of brain scans. More generally, we defer to the mere *trappings* of "science." This ready alienation of judgment presents an opportunity for all manner of cultural entrepreneurs who seek, not quite authority over others, perhaps, but rather to be the oracular *source* of such authority, whether in law, policy, psychiatry, or management. There is no arguing with a picture of a brain. Further, there is a ready market for the explanations offered by such entrepreneurs. Among those charged with the administration of human beings, there is a great hunger for scientific-looking accounts that can justify their interventions, as the aura of science imparts legitimacy to their efforts.

For example, professors of public policy dream of being able to use brain scans to predict a propensity, not only for violence, but also for tendencies like racial bias, as Jeffrey Rosen reported in the *New York Times*

Magazine in 2007. This would open a vista of social control previously only imagined, and expand the dominion of criminologists: if human behavior is electrochemically preordained, there remains no discernible ground on which to object to preemptive interventions directed against those identified as "hard-wired" malfeasants. Such interventions might take the form of surveillance, incarceration, or alteration (through drugs, surgery, or implants).

But neurolawyers and neurocriminologists are not exactly neuroscientists. The irony is that "we have no evidence whatsoever that activity in the brain is more predictive of things we care about in the courtroom than the behaviors that we correlate with brain function," according to Elizabeth Phelps, a cognitive neuroscientist at New York University, as quoted by Rosen. In other words, if you want to predict whether someone is going to break the law in the future, a picture of his brain is no better than a record of his past behavior. Indeed it is quite a bit worse, as the correlation of future behavior with brain abnormalities is weaker than it is with past behavior. Neuroscientist Michael Gazzaniga writes in his book *The Ethical Brain* that "most patients who suffer from…lesions involving the inferior orbital frontal lobe do not exhibit antisocial behavior of the type that would be noticed by the law." It is merely that people with such lesions have a higher incidence of such behavior than those without. So for the pragmatic purpose of predicting behavior, the story of neurological *causation* that is told by pointing to an image of a brain merely adds a layer of metaphysics, gratuitously inserted between past behavior and future behavior despite its lack of predictive power.

Rosen quotes Paul Root Wolpe, a professor of social psychiatry and psychiatric ethics at the University of Pennsylvania, as saying, "I work for NASA, and imagine how helpful it might be for NASA if it could scan your brain to discover whether you have a good enough spatial sense to be a pilot." But consider: NASA currently tests your spatial reasoning directly—the intellectual capacity itself, not a neurological *correlate* of it. This is done by putting you in a flight simulator and observing your performance in a pragmatic context similar to the one you would face as a pilot. But such a pragmatic orientation doesn't offer the excitement that comes with accessing a hidden realm of causation.

It may be worth recounting an episode from the history of science when hidden causes similarly had people excited. In the seventeenth century, one of the grand problems of science was to explain why things fall down. Descartes had developed a strictly mechanical, billiard-ball model wherein imperceptible particles impinging from above push things

downward. There were other, competing mechanical models. The problem was that no such mechanical picture could account for the findings of Galileo—namely, that bodies fall with a uniformly accelerated motion and the acceleration for all bodies is identical, regardless of size. This impasse surrounding what we now call gravity could be resolved by positing a force of attraction between bodies. Newton did just this, but in doing so he was attacked by the more doctrinaire "mechanicists," for whom it was a matter of principle that there could be no action at a distance. Newton was accused of re-introducing scholastic "occult qualities" into nature, precisely the kind of explanation that the mechanical philosophy set out to banish, just as the current reductionism in psychology wants to banish spooky notions like "soul."

While the mechanical philosophy confidently posited hidden mechanisms, on the assumption that there must be some cause that is similar to the ones we can see operating in the world, Newton was content to leave causes mysterious. He then proceeded to give a mathematical *description* of how bodies move under the mysterious attractive force: the inverse square law of gravity. Accepting the obscurity of gravity's causes seems to have freed Newton up to attend to the phenomena, and thus to accomplish his mathematization of the phenomena. The intransigently reductive position adopted by the mechanical philosophers was abandoned. It is worth noting that our understanding of gravity, though transformed by Einstein, remains agnostic on causes. Instead of spooky action at a distance, now we have even spookier distortions of space-time. Physicists seem to be less easily spooked than cognitive scientists.

The Autonomy of the Mental

Where does this leave us? I would like to make the case for giving due deference to ordinary human experience as the proper guide for understanding human beings. Such deference may be contrasted with the field of "neurophilosophy" (most famously, the work of Paul and Patricia Churchland), which is intent on replacing "folk terms"—such as "reflection" and "deliberation"—with terms that describe brain states. Needless to say, brain states are objective facts, whereas our introspective experience of our own mental life is inherently subjective. But this divide between the objective and subjective, between the brain and the mind, does not map neatly onto cause and effect, nor onto any clear distinction between a layer of reality that is somehow more fundamental and one that is merely epiphenomenal. For example, if you are told your mother has died, your

dismayed comprehension of the fact, which is a subjective mental event, will cause an objective physiological change in your brain.

In light of this causal power of the mental over the physical, we begin to wonder if it is right to think of these two types of reality as layered, in the sense that one is more causally effective than the other. It would follow from this doubt that re-describing our introspective experience of our own mental life in terms of brain states is optional, in this sense: the choice of description ought to depend on what you're trying to explain. Each description answers to a different sort of "explanatory request," issuing from different realms of practice. The "folk" description answers to the realm of everyday human experience, and the brain description answers to the realm of physiological investigation.

To insist that the brain description is superior to the mental one in a more comprehensive way, such that it may subsume the mental one and render it obsolete, there must be points of contact where the two descriptions conflict, so the better one can show itself as superior. This is what happened when, for example, the Copernican theory prevailed over the preceding view that the earth is the center of the universe. The problem with the neurological re-description of our mental life would seem to be that there is no such contact, hence no competition. Hence no reason for preferring one over the other, on any grounds other than pragmatic ones. There is an explanatory gap between our knowledge of the brain and what we know first-hand of ourselves, and it is difficult to imagine what kind of finding would bridge the gap. That there should be a neurological basis for our mental life is not controversial. But that beginning insight also seems to exhaust the contribution of brain scans to our self-understanding.

Bracketing the questions of the mind-body problem is unsatisfying. But such a lack of metaphysical satisfaction may be something we need to live with. To do so is a form of sobriety, as against the zeal of those who rush off to reform law, public policy, and ethics as though these ultimate questions had been settled, and always in such a way as to overturn what we know first-hand of our own agency.