NEURAL WORLDS AND REAL WORLDS

Patricia S. Churchland & Paul M. Churchland

Philosophy Department , UCSD

La Jolla CA 92093

ABTRACT

States of the brain represent states of the world. A puzzle arises when one learns that at least some of the mind/brain's internal representations, such as a sensation of heat or a sensation of red, do not genuinely resemble the external realities they allegedly represent: the mean kinetic energy of the molecules of the substance felt (temperature) and the mean electromagnetic reflectance profile of the seen object (color). The historical response has been to declare a distinction between objectively real properties, such as shape motion and mass, and merely subjective properties, such as heat, color and smell. This hypothesis leads to trouble. A challenge for cognitive neurobiology is to characterize, in suitably general terms, the nature of the relationship between brain models and the world modeled. We favor the hypothesis that brains develop high-dimensional maps whose internal relations correspond in varying degrees of fidelity to the enduring causal structure of the world. From this perspective, the basic epistemological relation is not "single-percept to single-external-feature" but rather "background-brain-maps to causal-domain-portrayed.

1. INTRODUCTION

Consider the following simplified theory: the brain constructs models -- of its body, of the world external to its body, and of some activities of the brain itself. By 'model', we loosely mean an organized representational scheme. The body is represented in its somatosensory and motor aspects, as well as in its inner milieu, including its drives, CO^2 levels, and blood pressure. Representation of the external world is probably anchored in a computational platform that is organized to predict the edibility, congeniality, hostility, and so forth of objects in space-time. Motor plans are run through emulators that simulate the body in its environment in order to predict the consequences of movements before their execution. [1] [2] [3] [4] [5] [6]

Some neural activities represent other neural activities as feelings, such as pain, fear, and fatigue; some represent sensory states as resulting from an encounter between an external object (e.g. a bee) and oneself (e.g. that stung me). According to Damasio[7], this fundamental representation of causal interactions between the external world and one's body anchors conscious self-representation; that is, the representation of self as inner versus world as outer. How much integration normally exists across various inner models and how integration, within and across models, is achieved, are open questions.[8] From the point of view of the user, of course, that the brain' constructs its distinctions between inner-me and outer-world is anything is but self-evident. The brain does not have introspective access to its world-modeling activities in the way that it has access to its pains, needs, and smells.

To a first approximation, this theory-cartoon is the conceptual framework that is assumed, implicitly or explicitly, by many cognitive scientists and neuroscientists. The point of outlining the theory is to make the main load-bearing pieces easily visible, as visibility aids exploration of what the framework entails, whether it constitutes a coherent package, what parts are radically incomplete, and whether some parts of the story might be entirely misconceived.

These are not the sorts of questions to be answered by a single, crucial experiment. Some might not be answerable at all at this early stage in the development of the brain sciences. Nevertheless, they are worth asking because background assumptions can turn out to be problematic. In addition, background assumptions, despite living a quiet life mostly in the background, do in fact motivate experimental research. They inevitably play a significant role in constructing and testing hypotheses. Finally, unless they are brought to the foreground occasionally, background assumptions tend to pass as obvious, god-given, beyond-question truths. This is not a good thing. Dogma is undesirable in any science, but especially so in a young field, such as neuroscience, where the fundamental principles that govern brain function and organization are only beginning to be understood.

2. PRIMARY AND SECONDARY QUALITIES

One matter that deserves attention concerns how we understand the relationship between the brain's models of the world and the world itself. In particular, what can we learn about the fidelity of the brain's representations relative to the things that are represented? An assortment of interconnected problems resides in this domain, and it will be useful to extract the most troublesome for further dissection.

According to conventional wisdom, some properties that are represented by the brain as in the world are not genuinely in the world at all, but are mere products of brain activity. The standard examples are colors, smells, and sounds. Thus, it will be argued, a peppermint-smell is what some brains create in response to particular molecules depolarizing receptor cells in the olfactory epithelium, while other brains are indifferent to those molecules. Although the molecules are in the external world, the smell itself is not. By contrast, the argument continues, certain properties, such as mass, motion and spatio-temporal relationships, are really in the world and our representations do resemble them. These are 'primary qualities', whereas smells and colors are merely 'secondary qualities' caused in nervous systems by primary qualities.

Galileo (1564-1642) was perhaps the first to postulate this distinction.[9] He aimed to give an explanation for the fact that way the world appears in experience may not be the way the world really is, and that we can discover how the world really is, despite how it may appear to us. Galileo's hypothesis was occasioned by the phenomenon of heat. He reasoned that, in reality, heat was probably the motion of tiny “corpuscles” (atoms, as conceived by Democritus [460-370 BC]). Galileo was struck by the fact that heat is not so apprehended by us in experience. The feeling of heat seems to have nothing whatever to do with perceived motion of anything. To explain why the appearance does not match the reality, he suggested that particle motion kicks off a causal process in the body that results in heat as we experience it. Given this and other differences, discovered by science, between appearance and reality, Galileo, and later John Locke (1632-1704), saw wisdom in distinguishing between real-world properties (primary qualities) and merely brain-constructed properties (secondary qualities). [10]

Something akin to the primary/secondary distinction is still commonly wheeled out in discussion of the conceptual framework outlined above. Despite its plausibility, the distinction carries the seeds of its own destruction.

Notice that the distinction privileges primary qualities, as uniquely real, by running very fast by a fact: there is no brain-independent or representation-free access to reality. If color and smell representations are the brain’s causal response to certain external stimuli, then so are spatial representations and motion representations. The brain cannot directly compare its representation of the external world with the external world itself, as one might compare the on-stage Wizard of Oz with the man behind the curtain.

To be sure, instruments can provide the brain with additional data, as Galileo's clever thermal-expansion thermometer provided him with data concerning heat. But instrumental data require human observation and theory-backed interpretation, both of which involve filtering through the lens of representational models. Hence "objective" instruments are not the general solution for funding a principled distinction between primary and secondary qualities.

Here then is the dilemma: if the distinctions between the inner-me and outer-world are made within the brain's representational models, then how does the distinction between brain-constructed properties and real-world properties get purchase? And why should we believe that primary qualities accurately characterize reality whereas secondary qualities are merely representational fabrications incident upon the brain's interactions with reality?

3. INTO IDEALISM AND OUT AGAIN

The philosopher Bishop Berkeley (1685-1753) realized that the arguments supporting the secondary qualities as mind creations starts us down a very slippery slope. (FIG.1) Once on the slope, we find we have slipped to the next stage: primary qualities, like secondary qualities, are nought but mind-created responses to a real world whose true nature we can never know . At the bottom of the slippery slope is the proposition that the so-called external world is, after all, nothing other than my idea of an external world. Ideas are the only things there really are. Furthermore, my apparently physical brain also must be nothing but an idea. In this case, only nonphysical minds -- only constellations of ideas -- genuinely exist. Classically, this view is known as idealism, and it is what awaits us at the bottom of the slope.

Berkeley gleefully took the trip to the bottom of the slope, and was an uncompromising idealist to the end.[11] In *The Critique of Pure Reason*, Immanuel Kant struggled mightily to stop two-thirds of the way down the slope, but hit bottom even so.[12] Georg Hegel (1770-1831) and other German Idealists in the nineteenth century found their intellectual home at the bottom. [13] Their efforts went into explaining the apparently 'physical' world in terms of the allegedly more basic world of mental ideas. Because some contemporary neuroscientists consider the primary/secondary distinction to be unavoidable and the slippery slope to idealism to be inevitable, we shall briefly discuss the disadvantages of idealism before returning to the question of the fidelity of brain-models to the world modeled.

A obvious and fundamental objection to Idealism is that it cannot account for the coherence and regularity in the (idea of the) external world or even in one's own mental life. Even simple regularities -- objects thrown in the air regularly fall towards the Earth, dry wood regularly burns, water reliably quenches fire – are inexplicable. Berkeley's solution appealed to the supernatural: God keeps the flux of ideas coherent. What happens in deep sleep or in coma when ideas vanish? What of the existence of the Universe before there were minds? Luckily, Berkeley's God saves the day by having all of that as ideas in His ample mind. The supernatural solution is transparently *ad hoc* -- a pixie-dust, magic-wand solution -- and Berkeley's contemporaries mercilessly tore it to shreds. Unfortunately for idealism, however, no one ever contrived a solution that was both intelligible and less *ad hoc* than Berkeley's.

Idealists have also been entirely unable to account for the progress of science in falsifying hypotheses (e.g. malaria is caused by bad air, the Earth is flat, the Earth is motionless). For Idealism, so-called falsification has nothing special to do with getting closer to the nature of reality. There are just ideas, and then some more ideas. As an explanation for scientific progress, Hegel, essentially a scientific naif, ground out famously obscure rigmarole about a dialectical process whereby the Universal Mind is inexorably coming to know Itself.

More generally, Idealism cannot begin to account for the spectacular developments in the sciences that falsified intuition-friendly ideas (such as that heat is a kind of fluid), and replaced them with explanatorily powerful but inobvious theories, such as the atomic theory, statistical mechanics, and evolutionary biology. Ironically, the very developments that provoked Galileo and Locke to seek epistemological ballast for the appearance/reality distinction are developments that stubbornly resist the Idealist's approach. Finally, and most awkward for the Idealist, it turns out that the brain sciences are making progress in explaining why things sometimes seem to be different from how they probably are; for example, why a straight pencil in water looks bent (FIG. 2), why two identical patches of gray appear to have different luminances [FIG. 3], and why a full moon looks much larger when seen on the horizon than when seen overhead. [14] In other words, the evidence increasingly supports the view that science can explain, in great and systematic detail, mental properties in terms of physical things (nervous systems), whereas the Idealists' program is left clawing at the air. [15] [16] [17]

Like any hypothesis, large-scale or small-scale, Idealism's value has to be measured in terms of its distinct explanatory and predictive results. On this criterion, Idealism scores in the hopeless range. It does no explanatory or predictive work in science or in ordinary life. Indeed, as Berkeley more or less admitted, even if one believes idealism to be true, one has no choice but to act as though it is false.

4. BRAIN-MODELS AND SURVIVING IN THE REAL WORLD

The epistemological question now before us is how productively to address the relationship between representational models and the world modeled. One response is to circumvent the whole mess with a pre-Galilean resolve to believe that we do in fact perceive reality precisely as it is. In other words, dodge the primary/secondary distinction and boldly claim there is a match between appearance and reality. Known as "naïve realism", this view is undone by its specious innocence regarding the developments of modern physics, astronomy, biology and chemistry. Additionally, it has to ignore the gathering evidence from cognitive science and neuroscience concerning the brain's constructive processes.

Consider for example the ambiguous figures, which can be seen as different objects even though the stimulus is unchanged (FIG 4); contours will be clearly seen where no luminance differences exist in the stimulus (FIG. 5); color will be perceived in regions which are actually white. Three-dimensional depth is regularly constructed from two-dimensional retinal arrays (FIG. 6). [18] [19]

Contrary to naïve realism, the conclusion that there is a perfect match between reality and representation is untenable. Nevertheless, a reasonable solution can be found by looking for less exacting and more flexible relationships, such as an informational correspondence between brains and the world.

Perhaps representational models in nervous systems are roughly map-like, in the sense that their internal abstract relationships map onto external relations between the various things in the world. [14] [20] The rough analogy is with a road map of a city where the real spatial relationships between actual roads are represented in the relationships between road-lines on the paper map. Just as road maps come in varying degrees of fidelity and detail, so brain models of the external world map the causal structure of the world with varying degrees of fidelity and detail. A frog's brain maps rather less of the causal structure of the world than a raven's brain; an infant's brain maps less of the causal structure of the world than an adult's brain; prescientific human brains map less of the causal structure of the world than scientifically trained brains. Note also that just as area maps are interest-relative in their composition, so the features mapped by brains are generally features that matter to the organism and to how it makes its living; i.e. 'me-relevant' features. Unlike two-dimensional paper maps, representation models in nervous systems will be multi-dimensional; indeed, probably very high dimensional. [5] [21]

Such coherence and predictive power as representational models enjoy is explained not by Berkeley's God, but by biological evolution. Animals are movers, and nervous systems earn their keep by servicing movement. Other things being equal (and there are a lot of other things), the better and faster the brain's predictive capacities relative to the animal's *modus vivendi*, the better the organism's behavioral portfolio in the cut-throat competition to survive and reproduce.

In the most broad terms, the solution found by evolution to the problem of prediction is to modify motor programs by sensory information. The value of the sensory impact is greater if it can signal me-relevant causal regularities between events. To achieve this, the system needs neural populations interposed between sensory receptors and motor neurons to find and embody higher-order regularities. The richer the interposed neuronal resources, the more sophisticated the statistical capacities and the greater the isomorphisms achievable between the brain's causal-maps and the world's causal relations. Importantly, much of the brain's input will be consequent upon the organism's own movements, exploratory and otherwise. This dynamic loop extracts vastly more information in a given time interval about the causal properties of the external world than could a purely passive system. In nervous systems generally, testing and having one's expectations met or surprised is the key to falsification and revision of representational models. [22] [23]

Simple predictive capacities are reflexes; prenatally organized predictive capacities are instincts; plasticity in predictive capacities permits long-term use of the results of trial and error in encounters with the world. Brains that represent certain higher-order regularities as allocentric objects enduring in space-time have a powerful representational tool for exploring and finding out yet more about the causal structure of the world. Nervous systems that can use external tools, such as microscopes and telescopes, extend their predictive capacities and expand the range of causal structures within their ken. Nervous systems that can invent and deploy theoretical tools, such as the notion of "valence" or "gravity" or "gene", extend that range even further. Organisms whose brains map knowledge sources in social groups can economize on what a single brain must know by exploiting the distribution of knowledge across individuals. [24] Organisms whose nervous systems enable them to deposit new knowledge into artifacts, offspring, and social institutions can revise and improve causal-maps across a few generations. [25]

Science, in its broadest sense, thus provides the general answer to the Galilean question: what epistemological grip can we get on the appearance/reality distinction? To provide a more detailed answer, we have to go through examples of scientific discovery. We can begin with homey examples, such as the calculations showing that the moon is not the size of a barn and as far away as a high cloud, or why neon color spreading is illusory, or why we think malaria is caused by the one-celled parasite, *Plasmodium falciparum,* transmitted by female mosquitoes as they penetrate the skin for a blood meal. We could move on to Lavoisier's experiments showing that in burning, so-called "de-phlogisticated air" (oxygen) rapidly combines with material in the wood. We could move next to the predictive and explanatory revolution wrought by Newton's Laws of Motion, and then on to Einstein's revolutionary updating in the Special Theory of Relativity. Learning a lot of science and a lot of history of science permits us to get a feel for the principles by which science moves forward in mapping the causal structure of the world.

There is no algorithm for making scientific progress, just as there is no algorithm for being rational. There are, however, instructive prototypes and useful rules of thumb: observe, think, test, don't be dogmatic, though don't change your mind too easily, don't get in a rut but don't give up too soon, take advantage of statistics but don't suppose good theories will simply waft up from statistical analyses, and so forth. [26] [27] [28] [29]

Ultimately, we want to understand in detail the neural mechanisms whereby common sense and science develop coherent representational models and what 'coherence' means in neural terms. We want to understand what the metaphor of 'isomorphisms-between-models-and-world' amounts to in nonmetaphorical terms.

To make progress on these questions, we shall rely increasingly on the entire range of brain sciences. We need greater understanding of the basic operations of neural networks, and vastly more knowledge concerning the evolution and the development of nervous system. Like the notion of a ‘map’, the notions of ‘representation’ and ‘information’ in nervous systems are poorly understood, and much more will have to be discovered to determine how they should be fleshed out and whether they should be replaced in favor of an explanatorily more powerful, if yet unimagined, framework. Our understanding of the nature of the computational processes that result in representational models, even quite simple models, is in its formative stages. Because nervous-system computation is dirty, me-relevant computation, as opposed to the clean, God’s-eye computation that is paradigmatic of electronic computers, we need fundamental, inventive, and biologically-sensitive new ideas. None of these lacunae is cause for despair, but rather an occasion for engaging in that bootstrapping business that is the scientific exploration of the causal structure of the world.

Acknowledgements

We are grateful for advice from F. Crick, A. Damasio, L. Goble, E. McAmis, S. Rickless

Websites for exploring visual phenomena and visual illusions:

Donald D. Hoffman:

http://www.aris.ss.uci.edu/cogsci/personnel/hoffman/Applets/index.html

Richard Gregory:

http://www.grand-illusions.com/index.html

Edward Adelson

http://www.

Dale Purves:

http://www.adm.duke.edu/alumni/purves

Al Seckel:

http://www.illusionWorks.com

Captions

Figure 1. The slippery slope to Idealism starts with the reasonable primary/secondary qualities distinction at the top. (Courtesy Marian Churchland)

Figure 2. The pencil appears to have a kink where it enters the water. This appearance is explained by the refractive properties of the water medium. (Courtesy D. Stack)

Figure 3. The squares marked A and B are in fact identical gray levels. B looks much lighter than A because the brain uses the assumption that the light source is behind the green pillar, and B is in its shadow. It therefore constructs the visual experience so that if were B out of the shadow, it would form a consistent part of the checkerboard pattern. (Courtesy Edward Adelson.)

Figure 4. Ambiguous figures. The figure at the top (A) can be seen either as a black tree-trunk object against a white background (as in B) or as a lumpy-edged white thing against a black background (as in C). (Rock 1975.)

Figure 5. On the left is a sequence of light blue lines on a plain white background. On the right, black extensions have been added to the ends of these same blue lines. Now the spaces between the lines appear to be blue, though in fact the

background is really white, and one sees a light blue worm with clear subjective boundaries. The subjective appearance of the color to the intervals between the lines in known as neon color spreading. (Courtesy Donald Hoffman)

Figure 6. Stereo (Courtesy Donald Hoffman)

1. Hoffman, D. D. *Visual Intelligence*. New York: Norton (1998).

2. Wolpert,, D. M., Ghahramani, Z. and Jordan, M. I. An internal model for sensorimotor integration. *Science* 269: 180-182 (1995).

3. Grush, R. Self, world and space: on the meaning and mechanisms of egocentric and allocentric spatial representation. *Brain and Mind* 1: 59-92 (2001).

4 Churchland, P. S. *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press (2002).

5 Churchland, P. M. *The Engine of Reason, The Seat of the Soul*. Cambridge, MA: MIT Press (1996).

6. Suri, R. E. and Schultz, W.

7. Damasio, A. R. *The Feeling of What Happens*. New York: Harcourt Brace (1999).

8. Gazzaniga, M. S. Cerebral specialization and interhemispheric communication. *Brain* 123: 1293-1326 (2000).

9. Galileo. From The Assayer. In *The Scientific Background to Modern Philosophy: Selected Readings* ed. by M. R. Matthews (New York: Hackett 1989). Galileo says, "But that there exists in fire, apart from shape, number, movement, penetration, and contact, some further quality which we call 'heat', I cannot believe. And I again judge that heat is altogether subjective, so that if the living, sensitive body be removed, what we call heat would be nothing but a simple word." (p. 60).

10. Locke, J. *An Essay Concerning Human Understanding* 1690. Modern edition ed. By P. H. Nidditch. Oxford: Oxford University Press (1975).

11. Berkeley, G. *Berkeley's Philosophical Writings*. Ed. By D. Armstrong. New York:Macmillan. (1965)

12. Kant, I.

13. Hegel, G. *Phenomenology of Spirit*. (1807)Trans. by A. V. Miller. (Oxford: Oxford Univ. Press. 1977).

15. Kaufman, L. and Rock, I. The moon illusion: I. *Science* 136: 953-961.

16. Palmer, S. E. *Vision Science*. Cambridge, MA: MIT Press. (1999).

17. Purves, D., Lotto, R. B., Williams, S. M., Nundy, S., and Yang, Z.. Why we see things the way we do: Evidence for a wholly empirical strategy of vision. *Phil. Trans. Royal Soc of London B* 356: 285-297(2001).

18. Ramachandran, V. S. Filling gaps in perception: Part II. Scotomas and phantom limbs. *Curr. Dir. In Psychol. Sci.* 2: 56-65 (1993).

19. von Helmholtz, H. L. F. *Treatise on Physiological Optics*. Trans. By J. Southal. New York: Dover (1925).

20. Gregory, R. L. *The Intelligent Eye*. New York: McGraw Hill (1970)

21. Shepard, R. N. and Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cog. Psych*. 1: 1-7 (1970).

22. Glymour, C. *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. (Cambridge, MA: MIT Press. 2001).

23. Schultz etc

24. Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*. Cambridge MA: MIT Press (1998)

25. Hutchins, E. *Cognition in the Wild*. Cambridge, MA: MIT Press. (1995)

26. Tomasello, M. *The Cultural Origins of Human Cognition*. Cambridge MA: Harvard Univ. Press. (1999).

27. Bacon, F. *Novum Organum* (1660). Modern edition: *The New Organon*. Edited by L. Jardine and M. Silverthorne. Cambridge: Cambridge University Press (2000).

28. Medawar, P. B. *Advice to a Young Scientist*. (New York: Basic Books 1981).

29. Crick, F. H. C. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books (1988)

30. Spirtes, P., Glymour, C. and R. Scheines. *Causation, Prediction, and Search*. New York: Springer-Verlag (1993).