

been noted by immunologists grappling with the origin of adaptive immunity (1, 31), are a consequence of similar selective pressures for diversification and co-evolution of recognition functions to retain affinity between interaction partners.

A hallmark of these specific recognition systems is that their genes are subject to intense diversifying selection. Large numbers of alleles are commonly found, and extraordinarily high levels of intraspecific polymorphism are typically achieved, in some cases resulting from accelerated rates of evolution (18, 32). Due to balancing selection, polymorphisms in these genes can persist for long periods of time and often predate species diversification. Trans-species polymorphisms have been described in the MHC (33) and in SI systems (34), and in both cases, divergence of some allelic lineages appears to have occurred at least 20 million years ago.

Another emerging commonality between recognition loci is their structural heteromorphism, which apparently reduces intralocus recombination events and prevents disruption of the co-adapted gene complex. The crucifer *S* locus has been extensively restructured by expansion or contraction of the physical distance between *SRK* and *SCR*, gene duplication, as well as rearrangement of these two genes relative to each other and to flanking markers (Fig. 1) (18, 35). Similarly, the MHC

has undergone frequent gene duplications and deletions during its evolution (33), and the mating-type locus of *Chlamydomonas* contains a highly rearranged region that causes suppression of recombination over a 1-megabase chromosomal region (36).

Thus, in many respects, the challenges facing research in the crucifer SI system are similar to those facing researchers of other recognition systems. Comparisons of these different systems should lead to insight into common selective pressures that drive the diversification and co-evolution of self/non-self recognition genes and shape the structure of their controlling loci.

References and Notes

1. F. M. Burnet, *Nature* **232**, 230 (1971).
2. D. De Nettancourt, *Incompatibility and Incongruity in Wild and Cultivated Plants* (Springer-Verlag, Berlin, 2001).
3. J. J. Rudd, V. E. Franklin-Tong, *New Phytol.* **151**, 7 (2001).
4. K. Ida et al., *J. Mol. Biol.* **314**, 103 (2001).
5. I. S. Nou, M. Watanabe, A. Isogai, K. Hinata, *Sex. Plant Reprod.* **6**, 79 (1993).
6. C. R. Schopfer, M. E. Nasrallah, J. B. Nasrallah, *Science* **286**, 1697 (1999).
7. H. Shiba et al., *Plant Physiol.* **125**, 2095 (2001).
8. T. Takasaki et al., *Nature* **403**, 913 (2000).
9. Y. Cui, Y.-M. Bi, N. Brugiere, M. Arnoldo, S. J. Rothstein, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3713 (2000).
10. A. P. Kachroo, C. R. Schopfer, M. E. Nasrallah, J. B. Nasrallah, *Science* **293**, 1824 (2001).
11. S. Takayama et al., *Nature* **413**, 534 (2001).
12. M. Kusaba et al., *Plant Cell* **13**, 627 (2001).
13. J. C. Stein, B. Howlett, D. C. Boyes, M. E. Nasrallah, J. B. Nasrallah, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8816 (1991).
14. S. Takayama et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1920 (2000).
15. S.-H. Shiu, A. B. Bleeker, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10763 (2001).
16. A. L. Hughes, *Cell. Mol. Life Sci.* **56**, 94 (1999).
17. V. Vanosthuyse, C. Miege, C. Dumas, J. M. Cock, *Plant Mol. Biol.* **46**, 17 (2001).
18. J. B. Nasrallah, *Curr. Opin. Plant Biol.* **3**, 368 (2000).
19. J.-L. Giranton, C. Dumas, J. M. Cock, T. Gaude, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3759 (2000).
20. D. Cabrilla, J. M. Cock, C. Dumas, T. Gaude, *Nature* **410**, 220 (2001).
21. S. L. Stone, M. Arnoldo, D. R. Goring, *Science* **26**, 1729 (1999).
22. C. Azevedo, M. J. Santos-Rosa, K. Shirasu, *Trends Plant Sci.* **6**, 354 (2001).
23. T. Nishio, M. Kusaba, *Ann. Bot.* **85** (Suppl. A), 141 (2000).
24. M. Watanabe et al., *FEBS Lett.* **473**, 139 (2000).
25. K. Hatakeyama et al., *Plant J.* **26**, 69 (2001).
26. M. Kusaba, C.-W. Tung, M. E. Nasrallah, J. B. Nasrallah, *Plant Physiol.* **128**, 17 (2002).
27. M. K. Uyenoyama, Y. Zhang, E. Newbigin, *Genetics* **157**, 1805 (2001).
28. D. P. Matton et al., *Plant Cell* **11**, 2087 (1999).
29. D. Charlesworth, *Curr. Biol.* **10**, R184 (2000).
30. A. J. Brown, L. A. Casselton, *Trends Genet.* **17**, 393 (2001).
31. J. Klein, *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
32. P. J. Ferris, C. Pavlovic, S. Fabry, U. W. Goodenough, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8634 (1997).
33. J. Klein, A. Sato, S. Nagl, C. O'Heuigen, *Annu. Rev. Ecol. Syst.* **29**, 1 (1998).
34. M. Uyenoyama, *Genetics* **139**, 975 (1995).
35. D. C. Boyes, M. E. Nasrallah, J. Vrebalov, J. B. Nasrallah, *Plant Cell* **9**, 237 (1997).
36. P. J. Ferris, E. V. Armbrust, U. W. Goodenough, *Genetics* **160**, 181 (2002).
37. J. B. Nasrallah, data not shown.
38. Supported by grants from the NIH, NSF, and USDA.

VIEWPOINT

Self-Representation in Nervous Systems

Patricia S. Churchland*

The brain's earliest self-representational capacities arose as evolution found neural network solutions for coordinating and regulating inner-body signals, thereby improving behavioral strategies. Additional flexibility in organizing coherent behavioral options emerges from neural models that represent some of the brain's inner states as states of its body, while representing other signals as perceptions of the external world. Brains manipulate inner models to predict the distinct consequences in the external world of distinct behavioral options. The self thus turns out to be identifiable not with a nonphysical soul, but rather with a set of representational capacities of the physical brain.

What Is "the Self"?

Descartes proposed that the self is not identical with one's body, or indeed, with any physical thing. Instead, he famously concluded that the essential self—the self one means when one thinks, "I exist"—is a nonphysical, conscious thing. At this stage of scientific development, the Cartesian approach is unsatisfactory for three reasons: (i) psychological

functions generally, including conscious thoughts such as "I exist," are activities of the physical brain (1, 2); (ii) aspects of self-regulation (e.g., inhibiting sexual inclinations), and self-cognition (e.g., knowing where I stand in my clan's dominance hierarchy), may be non-conscious (3); and (iii) as the Scottish philosopher David Hume (1711–1776) realized, there is in any case no introspective experience of the "self" as a distinct thing apart from the body (4). Introspection, Hume concluded, reveals only a continuously changing flux of visual perceptions, sounds, smells, emotions, memories, thoughts, feelings of fatigue, and so forth.

To identify the phenomenon that we want explained, it is useful to start with the idea that one's self-concept is a set of organizational tools for "coherencing" the brain's plans, decisions, and perceptions. Thus, if a brick falls on my foot, I know the pain is mine. I know without pausing to figure it out that "this body is my own," and that a decision to fight rather than flee is a decision affecting my body's painful encounter with the body of another. If I scold myself about jaywalking, I know that it is me talking to myself. We know that if we fail to plan for future contingencies, our future selves may suffer, and we care now about that future self. Sometimes we use "myself" to mean "my body," as when we say "I weighed myself." By contrast, when we say "I deceived myself," we are not referring to our physical bodies. We talk of our social and our private selves, of discovering and realizing ourselves, of self-control, self-improvement, and self-denial (5).

This remarkably diverse range of uses of the self-concept motivates recasting problems about "the self" in terms of self-representational capacities of the brain. Doing so deflates the temp-

Philosophy Department 0119, University of California, San Diego, La Jolla, CA 92093, USA.

*To whom correspondence should be addressed. E-mail pschurchland@ucsd.edu

tation to think of the self as a singular entity and encourages the idea that self-representing involves a plurality of functions, each having a range of shades, levels, and degrees. Further, it broadens the inquiry beyond humans to other species, suggesting that varying levels of coherencing operate in all nervous systems of any significant complexity. The reformulation also sets the stage for designing experiments to determine more precisely the types of self-representations nervous systems have, how they are connected to one another, and the nature of their neural substrates (6).

The expectation that the brain and behavioral sciences will eventually understand the nature of self-representational capacities is not universally shared. Traditionalists prefer to hive off the fundamental questions about the self or consciousness as philosophical in the “armchair-only” or “forever-beyond-science” senses of the term (7). The dominant ideology in academic philosophy, functionalism, acknowledges the relevance of the behavioral sciences, but discounts the neurosciences as largely irrelevant to making progress in understanding the higher functions (8). The functionalist rationale depends on an allegedly close analogy between psychological processes and running software on a computer. According to the analogy, the brain is only the hardware on which the cognitive software happens to run (9). The brain is thus deemed a mere implementation of the software. The corollary is that understanding the hardware is therefore unimportant, by and large, in figuring out the software. Though the analogy between cognitive functions and running software is not close but feeble, and though the corollary fails to follow, functionalism retains considerable popularity beyond the borders of neuroscience (10).

Mysticism and functionalism notwithstanding, questions about self-representation are steadily shifting into the province of the brain and cognitive sciences. This shift is part of a general trend enabled by the scientific advances in the 20th century at all levels of brain organization from synapses to systems. These advances, along with improvements in technology, data analysis, and computational modeling, have meant that virtually all topics concerning the mind are now vigorously explored at the interface of neuroscience, cognitive science, and philosophy. This has been the fortune, for example, of color perception (11), autobiographical memory (12, 13), the emotions (3, 14–16), decision-making (3, 12, 14, 17, 18), sleep and dreaming (19), and consciousness (6, 12, 15, 20).

As in any science, some discoveries force a more enlightened articulation of the very questions themselves. For example, the split-brain studies revealed that interrupting information flow between the two hemispheres by surgical section of the cerebral commissures gives rise to striking disconnection effects; that is, the perceptions and decisions of one

hemisphere are disconnected from those of its counterpart hemisphere (21). This implied that the “unity of self,” advertised by some philosophers as a “transcendental” necessity, was actually subject to anatomical manipulation. The results implied that such unity and coherence as exist in one’s self-conception depend not on transcendental necessity, whatever that might be, but on neuronal organization. A recharacterization of the phenomenon of unity of the self was consequently occasioned by the new empirical data.

A standard principle, illustrated by the split-brain results, is that the definition of the phenomenon to be explained *coevolves* with experimental discoveries. In the early stages of the scientific attack on any problem, accurate definition of the phenomenon is hampered precisely because not enough is known to permit an accurate definition. A pragmatic strategy is to begin by studying those cases agreed to be obvious examples of the phenomenon. Powered by this agreement, provisional, rough characterizations can leverage the science’s first stages, with refinements in the phenomenon’s definition emerging as the surrounding facts become clear. From a historical perspective, the interdependence of definition and discovery typifies the transformation of assorted problems of what was originally “pure” philosophy (e.g., the nature of fire, space, matter, life, the cosmos) into problems of the experimental sciences (2, 6).

Self-Representational Capacities

In the brain, some networks are involved in representing things in the external world, such as the face of Groucho Marx or a looming bus. Other networks represent states of the body, such as its posture or its need for water. Some networks operate on other representations, yielding meta-representations such as knowing that my need to flee is more urgent than my need for water, knowing that John dislikes me, or remembering that John hit me. Neural networks engaged in integrating such meta-representations are probably the ones most relevant to questions about self-representation.

Self-representations may be widely distributed across brain structures, coordinated only on an “as-needed” basis, and arranged in a loose and loopy hierarchy. We see the slow emergence and elaboration of self-representational capacities in children (22), and the tragic fading of these capacities in patients with dementia. Despite large gaps in our knowledge, human as well as animal studies have made it possible to begin to distinguish different types of self-representational functions, and in some instances, to identify, albeit in general terms, their neural dependencies.

Self-representational capacities include representing the internal milieu and viscera via chemical and neural pathways aimed largely at the brainstem and hypothalamus; representing

musculoskeletal structures via the somatic sensory system; representing autobiographical events via medial temporal lobe structures; deferring gratification and controlling impulses via prefrontal lobe and limbic structures; and representing the sequence of actions to take next, as well as representing where one is in space-time and the social order.

Studies of human patients reflect the multidimensionality of self-representation by showing that particular functions can be spared when others are impaired. For example, a subclass of amnesic subjects with bilateral lesions in the hippocampal and associated cortical structures are unable to acquire new knowledge and have lost essentially all autobiographical information. For example, the patient R.B. lives essentially within a moving 40-s time bin (3). Although R.B. does suffer diminished self-understanding, he nevertheless retains many elements of normal self capacities, including self-control in social situations and the fluent and correct use of “I.” He also knows his current body configuration and status, and he can engage in self-imagery, identify feelings such as happiness, and show sympathy with the distress of others. The existence of such amnesics is a counterexample to the seemingly obvious hypothesis that one’s self is constituted by personal narrative (9).

Schizophrenia, known to involve decreased prefrontal activity and increased striatal activity (23), presents a different dimension of self-dysfunction. During a florid episode, a schizophrenic may have good autobiographical memory, but suffer deep confusion about self/nonself boundaries, e.g., responding to a tactile stimulus by claiming that the sensation belongs to someone else or that it exists somewhere outside of him. Auditory hallucinations, often considered diagnostic of schizophrenia, exemplify integrative failure. The “voices” appear to be the patients’ own thoughts or inner speech, but they are not represented, and thus not recognized, as such (24, 25). The anesthetic ketamine and drugs such as LSD can trigger similar phenomena.

A patient with lesions in right parietal cortex, resulting in loss of sensation and movement on the left side of the body, may firmly deny that her left limbs are in fact hers. On occasion, a patient with limb denial will use the normal right arm to try to throw the paralyzed left leg out of the bed, insisting it is alien. Despite suffering compromised body-representation, the patients may nevertheless have normal autobiographical memory as well as other self-representational functions such as knowing whether they feel bored or hungry. Patients with lesions in the anterior cingulate region may exhibit alien hand syndrome. In these cases, the contralesional hand will sometimes behave as though it is independently controlled. Patients with alien hand syndrome sometimes control their embarrassing alien hand with verbal commands.

Self-regulating functions can also be selectively impaired. Lesions in prefrontal cortex, especially in the ventromedial region, have been followed by significant changes in self-control, and particularly in the capacity to inhibit unwise impulses, despite normal functioning of many other self-representational capacities. Personality changes commonly occur with prefrontal damage. Hitherto quiet and self-controlled, a person with lesions in the ventromedial region of frontal cortex is apt to be more reckless in decision-making, impaired in impulse control, and socially insensitive (3, 17, 18).

Evolution of Self-Representational Capacities

The most fundamental of the self-representational capacities probably arose as evolution stumbled on solutions for coordinating inner-body signals to generate survival-appropriate inner regulation. The basic coordination problems for all animals derive from the problem of what to do next. Pain signals should be coordinated with withdrawal, not with approach. Thirst signals should be coordinated with water-seeking, not with fleeing, unless a present threat takes higher priority. Homeostatic functions and the ability to switch between the different internal configuration for fight and flight from that needed for rest and digest require coordinated control of heart, lungs, viscera, liver, and adrenal medulla. Body-state signals have to be integrated, options evaluated, and choices made, since the organism needs to act as a coherent whole, not as a group of independent systems with competing interests.

The Neural Platform

The most basic level of inner coordination and regulation occurs in the brainstem, anchoring what Damasio refers to as "the protoself" (12). In vertebrates, the brainstem-hypothalamic axis is the site of convergence of signals from the viscera, internal milieu, and the somatic sensory system. Also located in the brainstem are nuclei that regulate vital functions, sleep-wakefulness cycles, arousal, attention, and the emotions. This level of integration, shared across many species, is the nonconscious neurobiological platform for higher levels of self-representation.

Other, more complicated and flexible aspects of the self demand greater computational resources. Wolpert (26) and Grush (27) have proposed that increased accuracy in planning and execution of movement in space-time is achieved by cortical models of the body in relation to its environment. Roughly, a somewhat sloppy inverse model is connected to an error-predicting forward model, and the two converge on a good answer to the problem of how to move a many-limbed body in just the right way at just the right time.

If, for example, the goal is to reach a plum, the inverse model gives a first-pass

answer to this question: What motor command should be issued to get my arm to contact the plum? Taking the command-proposal, the forward model calculates the error by running the command on a neuronal emulator, and the inverse model responds to the error signal with an upgraded command. Emulation is faster and safer than real-world feedback. Assuming the forward and inverse models are also capable of learning, this organization can be very efficient in acquiring a wide range of sensorimotor skills. With sufficient access to background knowledge, goal priorities, and current sensory information, emulators can compute accurate solutions to complex motor problems.

Rudimentary neuronal emulators, grounded in the basic coordinating and self-regulating functions, can in turn be upgraded to yield fancier inner models of planning. Emulators can facilitate making an appropriate movement after the target has become invisible, perhaps because the prey is in a cavity or the predator is sneaking up on the prey. More generally, with appropriate connectivity, an emulator could run off-line to plan for the long-term future, thus deploying extended body-image manipulation. Additional modification permits off-line emulation of cognitive states. For example, when planning the details of a raid, one may imagine oneself feeling anxiety while stalking the enemy camp, assessing the attentiveness of the camp guards, formulating specific intentions to outfox wary guards, and so on. Like body-image manipulation used in planning a climb, this is mind-image manipulation used in planning a complex, extended me-them encounter (27).

Consciousness and Self-Representation

An appealing hypothesis defended by Damasio (12) is that the self/nonself distinction, though originally designed to support coherencing, is ultimately responsible for consciousness. According to this view, a brain whose wiring enables it to distinguish between inner-world representations and outer-world representations and to build a meta-representational model of the relation between outer and inner entities is a brain enjoying some degree of consciousness. Thus, such a system could represent the relation between the thistle and itself as "that (outer) thing causes me (inner) pain." Conceivably, as wiring modifications enable increasingly sophisticated simulation and deliberation, the self-representational apparatus becomes correspondingly more elaborate, and therewith the self/not-self apparatus. On this hypothesis, the degrees or levels of conscious awareness are upgraded in tandem with the self-representational upgrades. Thus, chimpanzees, but not frogs, know whether they can be seen by a sub-

ordinate female but not the dominant male. Infant human development studies and non-human primate studies support these hypotheses (28, 29).

Whether neuroscience can build on these foundations to discover full and detailed explanations of all self-representational phenomena remains to be seen. Still, unpredictability obscures the destiny of essentially all neurobiological puzzles, including noncognitive functions such as thermoregulation. An abiding challenge in neuroscience is to discover the basic principles governing the integration of information at various levels of brain organization and at various time scales. This challenge is not confined to the neuroscience of self-representation, but confronts neuroscience generally.

References and Notes

1. P. M. Churchland, *Matter and Consciousness* (MIT Press, Cambridge, MA, ed. 2, 1988).
2. P. S. Churchland, *Neurophilosophy* (MIT Press, Cambridge, MA, 1986).
3. A. R. Damasio, *Descartes' Error* (Grossett/Putnam, New York, 1994).
4. D. Hume, *A Treatise of Human Nature* (1739); modern edition by L. A. Selby-Bigge, Ed. (Clarendon Press, Oxford, 1888).
5. G. Lakoff, M. Johnson, *Philosophy in the Flesh* (Basic Books, New York, 1999).
6. P. S. Churchland, *Brain-Wise: Studies in Neurophilosophy* (MIT Press, Cambridge, MA, in press).
7. See the essays by Z. Vendler and by C. McGinn, in *The Mind-Body Problem: A Guide to the Current Debate*, R. Warner, T. Szubka, Eds. (Blackwell, Oxford, 1994).
8. J. A. Fodor, *Synthese* 28, 97 (1974). For criticism of this view, see P. S. Churchland, T. J. Sejnowski, *The Computational Brain* (MIT Press, Cambridge, MA, 1992).
9. D. C. Dennett, *Consciousness Explained* (Little Brown, Boston, 1991).
10. S. Pinker, *How the Mind Works* (Norton, New York, 1997).
11. S. E. Palmer, *Vision Science* (MIT Press, Cambridge, MA, 1999).
12. A. R. Damasio, *The Feeling of What Happens* (Harcourt Brace, New York, 1999).
13. L. R. Squire, E. R. Kandel, *Memory: From Mind to Molecules* (Freeman, New York, 1999).
14. J. Le Douarin, *The Emotional Brain* (Simon & Schuster, New York, 1996).
15. R. R. Llinas, *I of the Vortex* (MIT Press, Cambridge, MA 2001).
16. J. Panksepp, *Affective Neuroscience* (Oxford Univ. Press, New York, 1998).
17. S. W. Anderson et al., *Nature Neurosci.* 2, 1032 (1999).
18. A. Bechara et al., *Science* 275, 1293 (1997).
19. A. R. Hobson, *Sleep* (Freeman, New York, 1989).
20. F. Crick, C. Koch, in *Problems in Systems Neuroscience*, J. L. van Hemmen, T. J. Sejnowski, Eds. (Oxford Univ. Press, Oxford, in press).
21. R. W. Sperry, *Science* 217, 1223 (1982).
22. S. Harter, *The Construction of the Self: A Developmental Perspective* (Guilford, New York, 1999).
23. A. Meyer-Lindenberg et al., *Nature Neurosci.* 5, 267 (2002).
24. C. D. Frith, *The Cognitive Neuroscience of Schizophrenia* (Erlbaum, Hillsdale, NJ, 1992).
25. G. L. Stephens, G. Graham, *When Self-Consciousness Breaks* (MIT Press, Cambridge, MA, 2000).
26. D. M. Wolpert et al., *Science* 269, 1880 (1995).
27. R. Grush, *Philos. Psychol.* 10, 5 (1997).
28. A. N. Schore, *Affect Regulation and the Origin of the Self* (Erlbaum, Hillsdale, NJ, 1994).
29. M. Tomasello, J. Call, *Primate Cognition* (Oxford Univ. Press, New York, 1997).
30. Special thanks to P. M. Churchland, M. Churchland, A. Churchland, J. Cohen, F. Crick, A. Damasio, D. Eagleman, and R. Grush.