nature neuroscience

Scanning the social brain

An ambitious neuroimaging study in this issue tackles the highly charged topic of race relations. Jennifer Richeson and colleagues at Dartmouth University previously showed that after white subjects interact with a black experimenter, they are impaired on a subsequent test of executive control (the well-known Stroop test); the degree of impairment is correlated with the subject's score on a test of implicit racial bias. The authors concluded that individuals with stronger racial bias must exert greater self-control when talking with people of other races, and this depletes their executive resources—corresponding to the everyday experience that a difficult mental task can drain one's ability to concentrate.

In the new study (page 1323), the authors used fMRI to examine the neural basis of this behavior and test the proposed mechanism. They find that cortical structures involved in executive control are activated in response to black faces, and that this activation is greater in individuals with stronger implicit racial bias. The most interesting point is that brain activation predicts cognitive impairment even better than does bias score, supporting the idea that depletion of executive resources is indeed responsible for the behavioral effect.

This study is likely to be widely discussed, so it is important to emphasize that it is not about biological determinants of racial prejudice. Although it describes neural correlates of implicit bias, it says nothing about the source of this bias; even more importantly, it says nothing about the relationship between implicit bias and actual behavior toward people of other races. The implicit association test (IAT) is experimentally convenient and widely used—readers can try it at www.tolerance.org —and IAT score has been correlated with (presumably fearful) amygdala responses to black faces[1]. But as discussed by William Gehring and colleagues in a News and Views (page 1241), how the IAT relates to real-world behavior remains controversial.

Despite this caveat, the paper is interesting because it exemplifies an important trend in cognitive neuroscience. Functional brain imaging is often criticized as descriptive, a new form of phrenology with no higher aim than attaching labels ('the racism center') to different patches of cortex. Admittedly the field has produced its share of descriptive papers, but the more sophisticated studies attempt to move beyond description, and to test mechanistic hypotheses.

'Mechanism' of course means different things to different people. To many biologists, it implies molecular analysis, but it seems implausible that human behavior can be reduced to molecular explanations, any more than quantum mechanics can explain molecular biology. The new study exemplifies a cognitive hypothesis: impaired executive performance after an interracial interaction results from temporary depletion of resources in brain areas that mediate executive behavior. To the extent that these parameters can be quantified and manipulated, one can formulate and test hypotheses that can be described as mechanistic, even if the underlying circuitry and firing patterns are not known. Direct tests would require manipulation of brain activity, which is difficult in humans, but cognitive models can still be strengthened by showing correlations between brain activity and behavioral performance. The Richeson study shows this for individual subjects, although the experimental design did not allow a test at the level of individual trials.

One intriguing example of trial-by-trial correlations is a recent study of the neural basis of fair/unfair economic judgments[2]. In the 'ultimatum game,' a sum of money is allocated to one player, the proposer, who must decide how much to keep and how much to offer to a second player, the responder. If the responder accepts the offer, the money is divided as agreed, but if he refuses, then neither player receives anything. Perhaps surprisingly, people routinely reject offers that they perceive as unfair, thereby sacrificing their own share to punish the selfish proposer. This behavior—baffling to economists—may have evolved to maintain altruism in groups by deterring selfish behavior[3].

The authors identified brain regions in responders that are activated in proportion to the unfairness of the offer; importantly, for a given level of unfairness (say, an 80/20 split), the strength of activation predicted whether the offer would be rejected on that trial. The relevant brain structures include the anterior insula, which is also activated by disgusting tastes and odors; the implication is that we are in some sense 'disgusted' by unfair behavior in other people and that we penalize them accordingly.

As these examples illustrate, neuroimaging is beginning to provide insights into a range of higher cognitive functions, including many for which there is no good animal model. Some of these studies also touch on areas of profound societal importance and controversy. In addition to race relations and economic justice, recent examples include perceived trustworthiness, moral reasoning, economic cooperation, social rejection and even consumer brand attachment.

The field of social neuroscience is still young, so many of its conclusions must be regarded as tentative. One obvious limitation is that scenarios tend to be highly simplified—video games instead of real high-stakes contests, faces on a screen instead of actual social interactions, and so on. This strategy has served well in other areas of biology—studying cells in monolayer cultures instead of *in vivo*, for example—and cognitive neuroscience will presumably advance as researchers devise ways (such as virtual reality displays) to make the experience more vivid and realistic. Other limitations include the frequent need to deceive subjects (with unknown degrees of success) and the common use of college student subjects, whose behavior and attitudes toward many societal issues may not be representative.

The ability to predict behavior from brain scans inevitably raises concerns about mind-reading and social control, but in reality this prospect is remote. Some researchers hope to develop practical applications ('neuromarketing' being one of the newest buzzwords; see for instance www.thoughtsciences.com), but for most of us, the main foreseeable benefit is the enlightenment that must surely come from a better understanding of our own mental processes.

1. Phelps, E.A. *et al. J. Cogn. Neurosci.* **12**, 729–738 (2000).
2. Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. *Science* **300**, 1755–1758 (2003).
3. Fehr, E. & Fischbacher, U. *Nature* **425**, 785–791 (2003).