Functional Neuroimaging Can Support Causal Claims about Brain Function

Matthew J. Weber and Sharon L. Thompson-Schill

Abstract

■ Cognitive neuroscientists habitually deny that functional neuroimaging can furnish causal information about the relationship between brain events and behavior. However, imaging studies do provide causal information about those relationships—although not causal certainty. Although popular portrayals of functional neuroimaging tend to attribute too much inferential power to the technique, we should restrain ourselves from ascribing it too little. ■

It is a healthy truism that secure inference in cognitive neuroscience requires converging methods. The brain is enormously complex, and vital knowledge about it has arisen at vastly different spatial and temporal scales, through observational methods as well as natural and engineered interventions. In light of the field's distinguished origins in lesion studies and single-unit neurophysiology, functional neuroimaging is too readily ceded pride of place in the canon of cognitive neuroscience methods—but that does not license its critics or its practitioners to invent limitations on the technique. Functional imaging is many things, but despite the familiar shibboleth, it is not merely correlational.[1] Imaging studies provide causal information, although not causal certainty, about the influence of brain activity on behavior. Moreover, the influence on brain activity of experimentally manipulated variables, such as stimuli and task instructions, is unarguably causal.

The popularity of brain stimulation techniques in cognitive neuroscience is rising, and researchers often justify using these techniques by invoking the inability of functional imaging to address causal matters (Iacoboni, 2009; Ruff et al., 2008; Wagner, Valero-Cabre, & Pascual-Leone, 2007; Uddin, Molnar-Szakacs, Zaidel, & Iacoboni, 2006). These justifications can be misleading. For example, Uddin et al. (2006) state that "functional magnetic resonance imaging (fMRI) provides only correlational information about the relationship between a given brain area and a particular cognitive task. Causal relationships between brain and behaviour can be tested with [transcranial magnetic stimulation (TMS)]." To say that "fMRI provides only correlational information" is to render inexplicable the utility of fMRI-guided TMS, which relies on the reasonable expectation that areas activated concurrently with task performance are likely to be causally involved in the task. On the other hand, the causal involvement of a given brain region in a task is no guarantee that the region implements the computation of interest in a given experiment; for example, diverse tasks across the domains of attention, memory, language, and cognitive control would suffer from well-targeted TMS to primary visual cortex.

We suspect that a key factor confusing this issue is the difference between correlational and merely observational information. When we observe two variables, neither of which is experimentally manipulated, we are powerless to assign causal influence. Some information from functional imaging studies has this character—for example, the relationship between brain activity and reaction times, preferences, or demographic variables. Likewise, because functional imaging studies do not manipulate brain activity directly, they cannot capture causality when the variables of interest are both neural. Miller, Vytlacil, Fegen, Pradhan, and D'Esposito (2010) show that disruption of lateral prefrontal cortex activity changes category selectivity in extrastriate cortex, a causal observation that could not be supported by fMRI alone. However, it is common to manipulate subjects' perception or behavior and measure brain activity at different levels of that manipulation. This experimental structure informs us about the influence of behavior on brain activity—which is not identical but unquestionably relevant to influence flowing the other way. Sarter, Berntson, and Cacioppo (1996) provide a useful framework with which to interpret the information gained from imaging experiments versus studies of lesions and brain stimulation: The latter inform us about P(B[ehavior] | A[ctivity]), whereas imaging studies testify to P(A | B). On this account, functional imaging's informativeness about causal hypotheses becomes clear: P(B | A) increases with P(A | B) via Bayes' rule.[2] Having construed conventionally "causal" and "correlational" methods in a symmetric probabilistic framework, Sarter et al. acknowledge "a fundamental asymmetry in the heuristic power" of these studies that

University of Pennsylvania

favors the former. We are content with this concession, merely adding that the advantage obtains only once we know where to stimulate or which patients to test.

One might object that the claim "functional imaging is merely correlational" does not deny these points—rather, it denies only that functional imaging by itself can provide certainty about the causal influence of brain activity on behavior. We doubt that the claim is often construed this narrowly, but thus read, it is of course accurate. However, it is plain that some functional imaging studies are specifically interested in the relationship between stimulus presentation or task-related behavior and brain activity. Atkinson and Adolphs (2005) claim that "functional imaging techniques support only correlational rather than causal claims about brain function," but they review several studies reporting that confronting participants with certain facial expressions elicits activity in amygdala, anterior insula, and striatum—a causal relationship. Several groups have shown that machine learning algorithms, applied to cortical activity evoked by visible lines and those masked into invisibility, can read out the orientation of those lines (Haynes & Rees, 2005; Kamitani & Tong, 2005); to claim that the brain activity they observed was not caused by the stimuli is to invalidate their work. Likewise, clinical studies sometimes seek to differentiate patient populations from one another or from healthy controls by observing differences in task-related brain activity. Fales et al. (2008, 2009) reported elevated amygdala activity in depressive patients viewing fear-related stimuli compared with controls in the same condition and separately that antidepressant medication normalizes prefrontal hypoactivity in depressives performing an emotional-interference task. To claim that the observed brain activity was not caused by the behavior is to negate the possibility of understanding the depressive brain as it interacts with the world in both the lay and the statistical senses to limit inference to the general tendency or main effect. However, these scientists have done better. There is no need to slight their contributions by making inaccurate disclaimers about their methods.

We acknowledge that, in general, the problem with functional imaging lists the other way. Both grant reviewers and the popular media are keenly interested in the brain–behavior relationships about which imaging research must remain tentative, and we often do not meet these enthusiasms with appropriate moderation. In addition, it is essential that cognitive neuroscientists be clear about the comparative strengths of nonimaging techniques. Still, to demean any of our methods cavalierly is to weaken the reputation of the whole field. Let us be forthright about our capabilities as well as our limitations.

Acknowledgments

The authors thank Mark D'Esposito and Joe Kable for insightful comments on earlier drafts of this manuscript. Brad Buchsbaum has published similar thoughts at http://flowbrain.blogspot.com/, although this article was conceived independently of his post.

Reprint requests should be sent to Matthew J. Weber, Center for Cognitive Neuroscience, University of Pennsylvania, 3720 Walnut Street Room B51, Philadelphia, PA 19104, or via e-mail: mweb@psych.upenn.edu.

Notes

1. Readers preoccupied with moral purity are encouraged not to look up one author's reference to "the correlational nature of functional neuroimaging methods" (Thompson-Schill et al., 2002).
2. Poldrack (2006) offers a fuller treatment of Bayesian inference from neuroimaging data.

REFERENCES

Atkinson, A. P., & Adolphs, R. (2005). Visual emotion perception: Mechanisms and processes. In L. Feldman-Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness*. New York: Guilford Press.

Fales, C. L., Barch, D. M., Rundle, M. M., Mintun, M. A., Mathews, J., Snyder, A. Z., et al. (2009). Antidepressant treatment normalizes hypoactivity in dorsolateral prefrontal cortex during emotional interference processing in major depression. *Journal of Affective Disorders, 112,* 206–211.

Fales, C. L., Barch, D. M., Rundle, M. M., Mintun, M. A., Snyder, A. Z., Cohen, J. D., et al. (2008). Altered emotional interference processing in affective and cognitive-control brain circuitry in major depression. *Biological Psychiatry, 63,* 377–384.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8,* 686–691.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology, 60,* 653–670.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685.

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2010). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience*.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10,* 59–63.

Ruff, C. C., Bestmann, S., Blankenburg, F., Bjoertomt, O., Josephs, O., Weiskopf, N., et al. (2008). Distinct causal influences of parietal versus frontal areas on human visual cortex: Evidence from concurrent TMS-fMRI. *Cerebral Cortex, 18,* 817–827.

Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience: Toward strong inference in attributing function to structure. *American Psychologist, 51,* 13–21.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., et al. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, and Behavioral Neuroscience, 2,* 109–120.

Uddin, L. Q., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2006). rTMS to the right inferior parietal lobule disrupts self-other discrimination. *Social Cognitive and Affective Neuroscience, 1,* 65–71.

Wagner, T., Valero-Cabre, A., & Pascual-Leone, A. (2007). Noninvasive human brain stimulation. *Annual Review of Biomedical Engineering, 9,* 527–565.