opinion piece

Functional MRI in Health Psychology and beyond: A call for caution

Tal Yarkoni
University of Colorado at Boulder

Functional magnetic resonance imaging (fMRI) has emerged as a primary tool in psychologists' arsenal. The ability to peer inside the waking brain as it produces language, perceives visual objects, and interacts with the social world has provided unprecedented opportunities to investigate and understand the neural architecture of human cognition. But as Spider-Man's uncle, Ben Parker, once famously said: with great power comes great responsibility. Because of the high cost of fMRI data collection, the opportunity cost of a poorly conducted fMRI study is liable to be much greater than that of the typical behavioral study. And because of the unusually strong influence brain images wield over the popular imagination (McCabe & Castel, 2008), it may be easier to convince the public, other scientists, and even one's self, of conclusions that are not wholly supported by the data.

In this short opinion piece, I lay out several reasons for exercising caution when conducting fMRI studies. Although the piece is intended for a health psychology audience, one of the points I hope to convey is that the challenges that face health psychologists using fMRI are very much the same ones that face other psychologists. So while the examples I'll use may have particular relevance to health psychologists, they should also illustrate much more general principles that apply to many, if not most, fMRI studies. This is by no means a comprehensive overview of the methodological and conceptual challenges involved in designing, analyzing, and reporting fMRI studies; I simply highlight a few issues that pose serious threats to the conclusions of many fMRI studies, and remain, in my view, widely underappreciated.

The vagaries of low power

When running an experimental study, it is desirable to ensure that the study is adequately powered; that is, if the targeted effect really exists in the population, the study will detect it with high probability. In practice, however, studies in most branches of psychology tend to be underpowered (Cohen, 1992), and fMRI studies appear to be particularly so (Yarkoni, 2009; Yarkoni & Braver, 2010). Because fMRI data acquisition is extremely expensive (typically several hundred dollars per hour), there's a strong pressure to collect as little data as possible. In practice, the modal fMRI sample size of 15 – 20 subjects often provides little power to detect anything but very large effects (Yarkoni, 2009). For example, a one-sample *t*-test performed on 20 subjects at a statistical threshold of $p < .001$ (the modal threshold in the fMRI literature) has only 40% power to detect even a canonically 'large' effect of $d = 0.8$. For a correlational analysis, the same sample size provides only 12% power to detect an extremely large correlation of $r = 0.5$. And yet simulations suggest that even a seemingly stringent (by behavioral psychology standards) threshold of $p < .001$ is insufficient to adequately control for false positives (Wager, Lindquist, & Kaplan, 2007). The inevitable conclusion is that the modal whole-brain fMRI analysis detects only a small minority of true effects while producing a high rate of false positives.

A related problem is that, when effects in underpowered studies *do* attain statistical significance, they tend to be grossly inflated (Yarkoni, 2009). The reason is that, when power is very low, the only way to detect an effect is to capitalize on chance. For instance, in a sample of 20 subjects tested at $p < .001$, the minimum statistically significant correlation is 0.67. A population correlation of, say, 0.3 will appear smaller or larger in any given sample due to sampling error; however, it will only be successfully detected in our small sample on those rare occasions when it is greatly inflated by chance. The problem is particularly acute in the context of the massive univariate analyses frequently performed in fMRI studies, because effects that may in truth be relatively weak and spatially diffuse will often appear to be spatially selective and extremely strong. For instance, if activity in half of the brain correlates 0.3 with some outcome variable in the population, we can expect the above sample to successfully detect the effect in fewer than 2% of voxels. And within the identified voxels, the observed correlation will be hugely inflated—averaging somewhere around 0.75 (Yarkoni, 2009). Paradoxically (and unfortunately), such biased findings may actually be easier to publish, because it's often more exciting to conclude that one has identified a highly circumscribed brain region that accounts for half of the variance in some outcome than to conclude that fully half of the brain is associated—but only weakly—with that outcome.

Not quite mind reading: the challenge of interpreting brain images

A second set of challenges concerns the interpretation of fMRI results. As difficult as behavioral results can be to interpret, neuroimaging results add an additional layer of complexity. Perhaps the most common approach to interpretation of fMRI results takes the following form: *we observed activation in region R; given prior literature demonstrating that R is involved in process P, this suggests that differences in process P, mediated by region R, may explain differences in outcome variable V.* This type of inference can be broken down into two strong claims: first, that there's a causal relationship between the observed changes in activation and some observed behavioral difference; and second, that we can readily infer what cognitive process such changes in activation reflects. In practice, both of these claims turn out to be surprisingly difficult to establish.

Consider the first claim. Suppose we observe, say, that the degree of right IFG activation in response to smoking cues predicts later success at abstaining from smoking. Can we conclude that IFG plays a *causal* role in mediating smoking abstinence? Not easily. Increased IFG activation in abstinent smokers could simply reflect the downstream effects of a critical upstream difference in a different process. For instance, smokers with greater motivation to quit might plausibly be more engaged with the task during scanning, and consequently expend more cognitive effort or spend more time attending to the on-screen stimuli. Because the blood-oxygen-level-dependent (BOLD) signal measured by fMRI sums approximately linearly over time, any increase in the amplitude or duration of neuronal processing will generally translate into a corresponding increase in the BOLD signal, irrespective of the efficacy of those processes in regulating behavior or other brain systems (Yarkoni, Barch, Gray, Conturo, & Braver, 2009). In other words, a change in IFG activation tells us only that there was *more* processing in IFG neurons; it doesn't tell us why. It certainly wouldn't imply that any cognitive process supported by IFG is the rate-limiting factor in ability to quit smoking. We can view the problem counterfactually: if we could manipulate smokers' brains to make them more

likely to quit, what systems would we target? Framed this way, it becomes clear that the mere presence of a correlation between regional changes in brain activity and some outcome variable provides little evidence of a direct causal relationship, because there are any number of other background processes where the critical causal locus could reside.

The second claim—i.e., that we can infer the cognitive processes involved in a task based on observed patterns of brain activity—is widely referred to as *reverse inference*, and is arguably still more problematic (Poldrack, 2006). The fundamental difficulty is in establishing *specific* mappings between cognitive states and activity patterns. To return to the above example, the right IFG is frequently implicated in emotion regulation and inhibitory control (Aron, Robbins, & Poldrack, 2004; Cunningham & Zelazo, 2007), so we might want to interpret our results as evidence that participants with greater inhibitory capacity are better able to regulate or inhibit their craving for cigarettes. But this type of claim, however intuitively compelling, is logically invalid. The fact that inhibitory control consistently elicits right IFG activity doesn't imply that right IFG activation is a *specific* marker of inhibitory control, because the same region could potentially also be activated by any number of other cognitive demands besides inhibitory control. Indeed, recent studies demonstrate that a distributed network of frontoparietal regions, including right IFG, is consistently and non-specifically activated by virtually all tasks involving goal-directed cognition (Duncan, 2010; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, submitted). Such findings suggest that, absent direct quantitative support, reverse inferences—currently a staple of Discussion sections in many articles—should be minimized or avoided. In practice, it is rarely possible to make strong quantitative statements about the causal processes implied by a particular pattern of brain activity.

A call for caution, not avoidance

The point of highlighting such concerns and limitations is not to suggest that fMRI has no place in health psychology and related fields; to the contrary, when used carefully, it can provide valuable information. Quite simply, if one's goal is to study the large-scale neural substrates of cognition and behavior, there are few better tools. Moreover, the challenges discussed above, while serious, are all solvable: low power can be addressed by increasing sample sizes, conducting hypothesis-driven tests, and performing multivariate analyses; causality can be established by complementing fMRI with other experimental techniques (e.g., TMS); and reverse inferences can be minimized or directly supported with quantitative estimates (e.g., Yarkoni et al., submitted).

What these concerns do hopefully underscore is that there are many ways for an fMRI study to fail—and that, paradoxically, it is often difficult to recognize that a study has failed precisely because the results appear more compelling than the underlying reality would dictate. When applied carefully, fMRI has the potential to facilitate better understanding of any number of health-related questions: who's likely to successfully quit smoking; how health-related messages influence people's thoughts and feelings; how people reason about risks and benefits related to their health; and so on. But achieving these aims requires us to address some difficult technical and practical challenges. It requires us to choose our sample sizes based on power calculations rather than convenience or budget; to weaken the inferences we draw from our data even if the net result is a less exciting manuscript; and to recognize that fMRI is only one tool among many, and is not a panacea for the many limitations of behavioral psychology. Provided health psychologists take such concerns

seriously, and exercise caution when using fMRI and other neuroimaging techniques, Uncle Ben will be pleased with us, and functional neuroimaging should have a bright future in health psychology.

References

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8(4)*, 170-7. doi: 10.1016/j.tics.2004.02.010.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112(1)*, 155-159.

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11(3)*, 97-104. doi: 10.1016/j.tics.2006.12.005.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14(4)*, 172-179. Elsevier.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: the effect of brain images on judgments of scientific reasoning. *Cognition, 107(1)*, 343-52. doi: 10.1016/j.cognition.2007.07.017.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data. *Trends in Cognitive Sciences, 10(2)*, 59-63.

Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience, 2(2)*, 150-158. Oxford University Press.

Yarkoni, T. (2009). Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power-Commentary on Vul et al. (2009). *Perspectives on Psychological Science, 4(3)*, 294-298. SAGE Publications. doi: 10.1111/j.1745-6924.2009.01127.x.

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., & Braver, T. S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis, 4(1). *Public Library of Science*.

Yarkoni, T., & Braver, T. S. (2010). Cognitive neuroscience approaches to individual differences in working memory and executive control: Conceptual and methodological issues. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition*. New York: Springer.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Manuscript submitted for publication*.

Tal Yarkoni
is a post-doctoral fellow at the Department of Psychology and Neuroscience, University of Colorado at Boulder

tal.yarkoni@colorado.edu