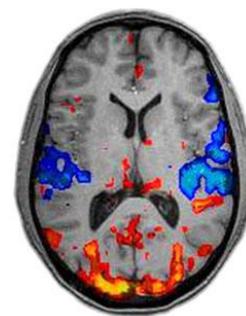


# What to Look for in an fMRI Paper

This brief guide is written for journalists and others who may not consider themselves fMRI experts, but would appreciate some advice about how to read an fMRI paper with an appropriate degree of informed skepticism. It is not meant to be prescriptive (or indeed exhaustive), but a few ideas of what to look for will hopefully benefit those wanting to report fMRI papers accurately in the media, as well as people who might simply wish to know how much they can reliably interpret from articles they read.



Like all science, fMRI is constantly evolving, so it is difficult to establish absolute standards that will be agreed by everybody. However, here are some pointers about common problems to look out for:

- The “neural correlates of x”. It’s not surprising that a particular cognitive function lights up in the brain somewhere – what else would you expect? Do the results tell us something new about how the brain works, or what cognitive processes might underlie the mental function of interest? Likewise, anything we do changes the brain, so a simple activity difference due to some intervention is not in itself interesting. Does the pattern of brain activity tell us something about *how* the intervention might work?
- The problem of “reverse inference”. If an fMRI paper links activity in one region with a single mental function, ask whether such a specific link is justified, as many brain regions are involved in multiple psychological processes and there is rarely a one-to-one mapping between activity in any brain region and a single mental state. The NeuroSynth web site (<http://www.neurosynth.org>) can be used to find out what functions are associated with any given brain region in the published literature.

*Case study: A NY Times op-ed suggested that we literally love our cell phones, based on a task in which participants viewed their phones, eliciting fMRI activity in the insular cortex, “which is associated with feelings of love and compassion.” But the insular cortex is active in as many as one third of all brain imaging studies, and is more often associated with negative than positive emotions.*

- Inappropriate significance thresholds. Is a statistical threshold used that corrects for the number of tests performed (e.g.,  $p < 0.05$  corrected), or at the very least a conventional uncorrected threshold (e.g.,  $p < 0.001$ )? An uncorrected analysis may contain false positive errors (activity that appears merely due to chance), and an unconventional threshold may have been selected to reveal only the activity the authors want you to see.
- Regions of interest (ROIs). If a ROI approach has been used (correcting for the number of voxels in specific regions, rather than the whole brain), it is essential the ROIs were selected independently of the analysis, which usually means based on the results of previous studies, or a different scan in the same experiment. Refer to the Introduction and whether the ROIs chosen seem plausible, or might they perhaps have been determined after the results were known but are presented as if they were a priori?
- Importance of control group. If the study compares two groups of subjects, or investigates an intervention, has an appropriate control group been chosen? Are control subjects matched on relevant variables (age, gender, etc)? In a treatment study, does the control group undergo the same testing schedule without the intervention? If not, the results may be uninterpretable.

*Case study: An fMRI study of children with dyslexia reported that after language training, activity in brain areas associated with language became more similar to that of typical-reading children. However, a control group who did not undergo the training was not included, meaning that brain activity might have changed over time anyway as a natural consequence of development.*

- How many subjects are involved? There's no perfect number, but anything less than 15-20 and you should ask serious questions about reliability of the results for most designs (although some studies in domains like perception that collect lots of data for each subject can use fewer). Were any subjects excluded after the data were collected – if so, why? If there are different conditions in the experiment, was their order counterbalanced to avoid possible order effects (e.g., due to fatigue or practice-related improvements)?
- Is there an interaction? Sometimes, authors observe that activity is present in one comparison and absent in another comparison, and erroneously conclude that there is a difference between the two effects. Similar conclusions might be drawn from seeing some activity in patients, but not controls (or vice versa). Such claims must be supported by a direct statistical comparison of the two effects (e.g., an interaction contrast), otherwise only limited inferences can be drawn.

*Case study: A study observed activity that was significant in the left hemisphere ( $p < 0.05$ ), and not significant in the right hemisphere ( $p > 0.05$ ), and concluded that their task specifically involves the left hemisphere. However, the right hemisphere effect may have just missed the threshold for significance (e.g.,  $p = 0.06$ ), in which case the difference between the two effects would be tiny and the claims of specificity unwarranted. An interaction contrast (e.g., is the left hemisphere effect significantly greater than the right hemisphere effect?) is necessary.*

- What does fMRI measure? Although fMRI signal is related to neural activity, exactly how is unresolved. Be careful of claims from fMRI data about what individual neurons might be doing. Avoid calling a behaviour “hard wired” on the basis of fMRI data and, because such data are correlational, avoid writing that a region “causes” or is “necessary” for a process.
- “Mind reading” and decoding. Various statistical approaches are being increasingly used to evaluate the amount of information contained in fMRI data. For example, it can be possible to determine from patterns of fMRI activity the category of object a participant is viewing. Although these studies provide useful information about the type of information being represented, they do not constitute “mind reading” in the way it is generally understood. Using that and similar terms can cause substantial confusion.

Despite the seductive allure of blobs on brains, no single fMRI study provides the final “answer” to anything in isolation, so be very skeptical of bold claims in journal articles or press releases. Are the results described fully and accurately, mentioning limitations and caveats, or are authors selective in what they focus on? Can their claims be justified based on the data? Ask whether you would have drawn the same conclusions based on the results. Include appropriate caveats in your reporting, and resist the inclusion of hype or exaggeration.

Try to obtain a quote from an independent expert before running with unverified assessments of a study's importance from the authors themselves or their university press releases. Open availability of the data can help with independently verifying the findings.

Many fMRI scientists will be happy to answer questions or fact check copy if it would be helpful. Do ask!

The list maintained by Micah Allen (@neuroconscience) includes a number of fMRI experts: <http://twitter.com/neuroconscience/lists/cogneuro/members>

Created by: Jon Simons (@js\_simons), Russ Poldrack (@russpoldrack), Ed Yong (@edyong209), with helpful contributions from Jonathan Peelle (@jpeelle), Dan Lurie (@danjlurie), Michael Waskom (@michaelwaskom), Beau Sievers (@beausievers), and others. Updated Dec 2014.