Are Cognitive Functions Localizable?

The Fall 2011 issue of this journal published a two-paper section on “Neuroeconomics.” One paper, by Ernst Fehr and Antonio Rangel, clearly and concisely summarized a small part of the fast-growing literature. The second paper, “It’s about Space, It’s about Time, Neuroeconomics, and the Brain Sublime,” by Marieke van Rooij and Guy Van Orden, is beautifully written and enjoyable to read, but misleading in many critical ways. A number of economists and neuroscientists working at the intersection of the two fields shared our reaction and have signed this letter, as shown below. Some of the paper’s descriptions of empirical findings and methods in neuroeconomics are incomplete, badly out of date, or flatly wrong. In studies the authors describe in detail, their skeptical interpretations have often been refuted. In a variety of tasks, the authors describe the strong spatial modularity view. But $15$ years is a long time ago in neuroscience. Skepticism about spatial localization is not new. Van Orden and Paap (1997) also attacked the strong spatial modularity view. But it is certainly true that a single region of the brain identified by one study is likely to be active in a variety of tasks. However, van Rooij and Van Orden substantially overrepresent the case against functional localization in their Appendix by only listing a highly selective set of studies with the widest range of different interpretations.

Van Rooij and Van Orden did not describe the many methods that are actively used now to check whether localized spatial regions predict common functional activity across tasks. For example, cross-method studies combine fMRI, causal manipulation of activity in targeted regions of the brain (using transcranial stimulation with magnetism or direct current), and behavior of patients with focal lesion damage in certain regions. Results from these types of studies will simply not fit together if there is no functionally reliable localization. Another tool is “activation likelihood estimation,” a meta-analytic method of combining results from many different studies. Regions that appear in tasks with a common functional component repeatedly are picked out by this approach, and regions with study-specific activity disappear. This method has been used since Turkeltaub, Eden, Jones, and Zeffiro (2002); the latest version is described in Eickhoff, Bzdok, Laird, Kurth, and Fox (2012). Yarkoni, Poldrack, Nichols, Van Essen, and Wager (2011) offer other methods of computational meta-analysis of neuroimaging data.

Van Rooij and Van Orden offer two specific examples in this part of their discussion: neuroimaging studies of ultimatum bargaining and trust games, ability, emotional responses, aesthetic judgments, and abstract thinking. In addition, there are many examples of fairly localized functional specialization. Examples include language areas and pathways, somatosensory cortex (for example, perception of touch), face recognition, and visual areas clearly corresponding to distinct steps in visual processing. For higher-order cognition, it is certainly true that a single region of the brain identified by one study is likely to be active in a variety of tasks. However, van Rooij and Van Orden substantially overrepresent the case against functional localization in their Appendix by only listing a highly selective set of studies with the widest range of different interpretations.

In the first part of their paper, van Rooij and Van Orden argue that neuroimaging studies are based on a faulty model of how the brain works because brain functions cannot be spatially localized to particular brain regions or networks. Skepticism about spatial localization is not new. Fifteen years ago, Van Orden and Paap (1997) also attacked the strong spatial modularity view. But it is certainly true that a single region of the brain identified by one study is likely to be active in a variety of tasks. However, van Rooij and Van Orden substantially overrepresent the case against functional localization in their Appendix by only listing a highly selective set of studies with the widest range of different interpretations.

In addition, there are many examples of fairly localized functional specialization. Examples include language areas and pathways, somatosensory cortex (for example, perception of touch), face recognition, and visual areas clearly corresponding to distinct steps in visual processing. For higher-order cognition, it is certainly true that a single region of the brain identified by one study is likely to be active in a variety of tasks. However, van Rooij and Van Orden substantially overrepresent the case against functional localization in their Appendix by only listing a highly selective set of studies with the widest range of different interpretations.

Van Rooij and Van Orden did not describe the many methods that are actively used now to check whether localized spatial regions predict common functional activity across tasks. For example, cross-method studies combine fMRI, causal manipulation of activity in targeted regions of the brain (using transcranial stimulation with magnetism or direct current), and behavior of patients with focal lesion damage in certain regions. Results from these types of studies will simply not fit together if there is no functionally reliable localization. Another tool is “activation likelihood estimation,” a meta-analytic method of combining results from many different studies. Regions that appear in tasks with a common functional component repeatedly are picked out by this approach, and regions with study-specific activity disappear. This method has been used since Turkeltaub, Eden, Jones, and Zeffiro (2002); the latest version is described in Eickhoff, Bzdok, Laird, Kurth, and Fox (2012). Yarkoni, Poldrack, Nichols, Van Essen, and Wager (2011) offer other methods of computational meta-analysis of neuroimaging data.

Van Rooij and Van Orden offer two specific examples in this part of their discussion: neuroimaging studies of ultimatum bargaining and trust games,
and of the ambiguity-risk distinction. Both discussions are misleading.

Sanfey et al. (2003) first used fMRI during ultimatum bargaining. Van Rooij and Van Orden note that:

If we take the results concerning fairness and generosity from the trust game, together with the previous results concerning unfairness and punishment from the ultimatum game, it would appear that responses to fairness and unfairness are formulated in different parts of the brain.

However, recent studies corroborate many of the conclusions of Sanfey et al. (2003) quite well. For instance, Chang, Smith, Dufwenberg, and Sanfey (2011) find substantial overlap between the neural systems involved in decision-making in the trust game and the regions identified by Sanfey et al. in the ultimatum game. Furthermore, if Sanfey et al. (2003) had wrongly interpreted the roles of the brain regions identified in their study, then disrupting activity in those brain regions using transcranial magnetic stimulation should have no effect on responder behavior in ultimatum games. But it does, as Fehr and Rangel point out (citing Knoch, Pascual-Leone, Meyer, Treyster, and Fehr 2006).

In their example of the neural distinction between responses to risky and ambiguous gambles, van Rooij and Van Orden compare results from Smith, Dickhaut, McCabe, and Pardo (2002) and Hsu, Bhatt, Adolphs, Tranel, and Camerer (2005). But these papers are not comparable, because the tasks are quite distinct. Smith et al. use variants of Ellsberg colored-ball tasks, choosing between an ambiguous gamble and a risky gamble (in some trials). Hsu et al. had subjects choose between certain amounts or individual gambles. Thus, Smith’s contrasts do not directly measure the difference between computation of ambiguous gamble valuation compared to risky gamble valuation; Hsu et al.’s analyses do measure that difference (using a conjunction of activity in all three tasks).

Moreover, the Hsu et al. (2005) finding of stronger lateral activity in the orbitofrontal cortex in response to ambiguity has been corroborated in two ways. In their original paper, they predict that people with brain damage in that area of the brain would be ambiguity-neutral, and they test and confirm this hypothesis in their paper. In addition, their findings were closely corroborated by Levy, Snell, Nelson, Rustichini, and Glimcher (2010, fig. S5). Van Rooij and Van Orden do not mention either of these corroborations.

In addition to their limited discussion of the neuroeconomic literature, van Rooij and VanOrden misrepresent the statistical methodology of neuroimaging. For example, they write that “the spatial approach to studying the brain assumes that the brain can be treated as the sum of its parts...this approach underlies what is often called the General Linear Model of the brain.” There is no so-called “General Linear Model of the Brain.” The phrase “General Linear Model” in the context of neuroimaging refers simply to the statistical technique of multiple linear regression. The term General Linear Model is never used by neuroscientists to describe a model of how the whole brain works; for example, the term does not appear in any widely used neuroimaging textbook.

Another misrepresentation involves the discussion of multiple tests of statistical hypotheses. Van Rooij and Van Orden say “any contrast using brain images can be counted on to make ‘discoveries’.” Neuroimaging researchers are well aware of the potential for false positives. Every widely used neuroimaging textbook contains detailed discussions of the multiple comparisons problem and methods for addressing it (for example, Ashby, 2011, chap. 6; Poldrack, Mumford, and Nichols, 2011, chap. 7; Huettel, Song, and McCarthy, 2009, chap. 12). The best standard of practice in neuroimaging research is to describe and account for the multiple testing problem (Poldrack, Fletcher, Henson, Worsley, Brett, and Nichols 2008).

To summarize what you have just read: the criticism that spatial identification of brain regions and circuits cannot be identified with functions is an old criticism. It has largely been disproven, is no longer widely believed in neuroscience, can be tested with various methods (cross-method and meta-analysis), is largely disproven for one example they discuss (ultimatum games) by newer studies, and is disproven for the other example (ambiguity) by data in the same paper they cite.

More broadly, the description of the general methodology of neuroeconomic research by van Rooij and Van Orden is misleading. Even given that articles in this journal are not meant to be comprehensive literature reviews, their sourcing on general neuroeconomics is surprisingly thin, ignoring the leading edited compilation (Glimcher, Camerer, Fehr, and Poldrack 2009), several reviews (for example, Fehr and Camerer 2007; Loewenstein, Rick, and Cohen 2008), recent papers from inside economics (such as Bernheim 2009; Rustichini 2009), and a thoughtful recent book (Glimcher, 2011). In their online Appendix, they strangely, but clearly, misplotted many brain areas.

The more interesting part of the van Rooij and Van Orden paper, which is closely linked to Van Orden’s own research, is about why attention to the detailed time course of neural activity is important (as a supplement to spatial understanding). There is no disagreement here, since virtually all neuroscientists who use methods with good time resolution do so to understand fine-grained (sub-second) temporal dynamics, neuroeconomists included. Indeed, the Fehr and Rangel paper in the same issue of this journal discusses drift diffusion
models, which predict choices, response times, and other temporal features.

For more than 10 years neuroeconomists have thought about the criticisms and methodological imperfections described by Van Orden and Papp (1997) and have figured out how to respond to those criticisms. It is unfortunate that the paper by van Rooij and Van Orden does not reflect this progress, and therefore misleads readers about the state-of-the-art, rather than educating them.

Colin Camerer, Caltech
Alec Smith, Caltech

Signers (ordered by timing of reply):
Camelia M. Kuhnen, Northwestern
Donald T. Wargo, Temple
Gregory Samanez-Larkin, Vanderbilt
Read Montague, Virginia Tech
Dino J. Levy, NYU
David Smith, Duke
Dar Meshi, Freie Universität Berlin
Peter H. Kenning, Zeppelin University, Germany
John Clithero, Caltech
Bernd Weber, Bonn
Todd Hare, Zurich
Scott Huettel, Duke
Camilla Josephson, Linköping University
Mathieu d’Acremont, Caltech
Daria Knoch, Basel
Ian Krajbich, Zurich
Benedetto De Martino, University College London
Peter N. C. Mohr, Freie Universität Berlin
Jan Barton, Emory
Marja-Liisa Halko, Aalto University
Christina F. Chick, Cornell University
Lorena Gianotti, Basel
Hauke R. Heekeren, Freie Universität Berlin

References


assumptions: 1) “brain instan-

All these statistical operations are subsumed by
what is commonly known as the General Linear
Model, a statistical model that includes regression
and related methods such as analysis of variance.
These methods form the basis of statistical param-
metric mapping, for example (for example, Friston,
Holmes, Worsley, Poline, Frith, and Frackowiak
1994; Friston et al. 2007). The approach is widely
used and cited in the neuroimaging literature (for
example, Mumford and Nichols 2009). The stan-
dard regression model states that an observable
equals a weighted sum of predictors, plus error.
Relying on these techniques in imaging analyses
is logically equivalent to adopting a general lin-
ear model of the brain. As our commenters write,
neuroscientists do not use the phrase “general lin-
ear model of the brain,” because they do not view it
as an accurate description of how the brain works.
Knowingly or not, neuroimagers using the sub-
tractive logic assume, as van Rooij and Van Orden
wrote in the original article, “that the brain can be
treated as the sum of its parts.”

According to Roskies’s (2010) defense of subtrac-
tion techniques, they rely on three crucial, a priori
(before the fact), assumptions: 1) “brain instanti-
tates mind,” 2) “different localized regions of
tissue have different and stable functionalities,” and
3) “blood flow is a guide to neural activity.” Each of
these assumptions is questionable.

Regarding assumption 1, debate on the nature of
mind-body relations is a historically persistent
and unsettled topic in biology, psychology, and
metaphysics.

Assumption 2 that “different localized regions of
tissue have different and stable functionalities”
is best described as a working research hypo-
thesis: If the brain is composed of distinct, modular,
functional regions that are causally transparent to
thought and behavior, then differences in spatial
brain activation can be functionally interpreted.
The potential difficulties with this approach have
long been recognized, and remain conspicuously
unanswered by advocates of subtraction such as
Camerer, Smith, and their co-signers.

Our concern is that much of the neuroimaging
enterprise is recapitulating previously established
localizationist failures in behavioral methods. In
1896, F. C. Donders, a Dutch ophthalmologist,
developed the hypothesis of distinct component
sub-operations intervening between a stimulus and
response, and devised the subtractive method to
investigate these components. The limitation of
Donders’s method, which persists to this day, is that
the hypothesized distinct components and subtasks
must be known before the fact for the resulting
subtractions to be meaningful (Uttal 2001). For
instance, lesions to certain parts of the brain, and
how they may affect specific behaviors, were among
the earliest localizationist targets in behavioral
studies (for example, Broca 1861). Nevertheless,
the storied history of language-based dissociation
studies has failed to converge on a view of the brain
as including fixed modules or even on criteria
for determining if such modules are an accurate description of how the brain works (Van Orden, Pennington, and Stone 2001).

The subtraction approach to studying localization of brain activity depends crucially and circularly on both theory and task. As the comment from Camerer and Smith illustrates, the theoretical debates inevitably degenerate into irresolvable disputes about what aspects of task and method are displayed in a specific hypothetical brain function (Van Orden, Pennington, and Stone 2001). Given a vast and variegated set of potential brain studies from which to choose, and the fact that minor task details routinely yield contradictory theoretical narratives, it is unsurprising that they or we or anyone else could reach endlessly contradictory conclusions regarding the same basic human activity, such as gambling.

Camerer, Smith, and their co-signers discuss how cross-method studies seek to resolve such issues. In practice, cross-method studies only amplify the already vast pool of potential contrasts, priorities, and perspectives on measured variables. They also claim that meta-analysis provides a solution, but meta-analysis is rooted in the self-same linear statistical system: “[T]he null hypothesis is that the n peak coordinates reported in the set of studies to be analyzed are randomly and uniformly distributed throughout gray matter” (Wager, Lindquist, and Kaplan 2007, p. 153, italics added).

By the mid-twentieth century, Saul Sternberg (1969) proposed an alternative method to discover subtask components in behavioral data that relieved the requirement that scientists know, before the fact, the functional components that intervene between a presented stimulus and participant’s response. Sternberg proposed an additive factors logic. Broadly speaking, this approach used factorial manipulations to examine how components might combine their effects additively.

Sternberg’s approach was potentially tractable, and within a few decades, an enormous set of psychological factors were assessed in multifactor behavioral experiments. Regarding the localizationist enterprise, to date, “not one cognitive mechanism exists on which cognitive scientists can agree about its boundaries, its empirical shape, or details about its function” (Van Orden, Holden, and Turvey 2005, p. 121). Instead, complex, contextually embedded chains of interactions among factors are routinely observed across studies (Van Orden, Pennington, and Stone 2001). This outcome is expected in the absence of modularity.

Indeed, Camerer, Smith, and their co-signers implicitly arrive at the same conclusion when they refer to the localization of “multiple regions” and “brain networks.” The imaging literature now reports increasingly complex and distributed brain networks, measured during both task and untasked conditions (for example, Bullmore and Sporns 2009; Ciuciu, Varoquaux, Abry, Sadaghiani, and Kleinschmidt 2012). Multiple motifs, or repeating connectivity patterns, are nested within a network’s connectivity. Different architectures that combine these low-level motifs provide for distinct local network flow patterns. The emerging emphasis is functional connectivity, the study of such patterns is sometimes called “connectomics,” by analogy to genomics. In stark contrast to subtracted images that identify just a few functional locations, the network approach reveals so many interconnected regions, scientists must wonder what regions are not functionally associated with a targeted activity (for example, Anderson 2010).

Assumption 3, that “blood flow is a guide to neural activity,” means that neuroimagers typically highlight those brain regions where the relative difference in blood flow between baseline and experimental sessions are at or near maximum. But what is special about a brain region that displays the largest differences in nutrient consumption?

Instead of considering blood flow measures in a brain, consider for a moment traffic flow measures in a city. Imagine that a researcher has data available on vehicles operating with the highest fuel consumption at any given time (that is, maximum nutrient consumption), or vehicles with full gas tanks (that is, maximum Hb to dHb ratios). It would clearly be misguided to conclude that only these vehicles are participating in transportation activities. In reality, the overwhelming bulk of transportation activities in a crowded city corresponds to vehicles that are not at extremes of fuel use or fuel storage—even stopped vehicles, at intersections for instance, crucially support transportation. It seems similarly unwise to assume that the brain areas with high nutrient consumption or BOLD signals are the only parts of the brain involved in a specific act of thinking.

As stated, fMRI statistical practices are rooted in the General Linear Model. Linear statistical methods are designed to discover and distinguish separable sets of constants (or means) shrouded by unsystematic, independent, and homogeneous sources of noise. As the fractal physiology literature cited by van Rooij and Van Orden makes clear, physiological signals rarely conform to such static assumptions. Instead, strong autocorrelation and inherent (fractal) fluctuations are fundamental properties of nearly all physiological signals. Long ago, it was recognized that applying static linear statistical methods to measurements that express intrinsic, long-range fluctuations, such as cardiovascular diffusion and flow, can be counted on to routinely yield spurious differences (as an example, see the Yule-Slutsky effect discussed in Klein, 1997).

The patterns we are discussing are also referred to as “scale-free.” In a model of the brain using subtractive logic, in which different tasks are associated with specific locations in the brain, the brain is...
divided into modules that can be labeled as either contiguous or isolated based on their distance from each other. In a scale-free system, there is no “typical” distance or time scale, in the sense that any chosen size of scale leads to different conclusions about which modules are involved and interconnected. Scale-free behavior is routinely reported in neuronal, fMRI BOLD, EEG, and behavioral signals. On this point, imaging and behavioral studies are in an unusual position of agreement regarding the empirical facts. Ubiquitous scale-free patterns in physiology and behavior implicate underlying flexible, self-organizing, and dynamic pattern formation processes, as discussed by van Rooij and Van Orden.

This exquisite dynamic flexibility has crucial implications. The instantaneous neurophysiologic details of every human thought or act may be sufficiently idiosyncratic that neuroimages reveal largely transient patterns. If so, then imaging practices that seek to determine a specific location in the brain associated with a specific cognitive function, are analogous to attempting to infer a theory of lightning by tracing the paths of reams of individual lightning bolt images. As Uttal (2001) famously claimed, subtractive brain imaging is effectively a new phrenology.

Marieke van Rooij
John G. Holden
University of Cincinnati
Cincinnati, Ohio

*Note: Guy Van Orden passed away May 11th, 2012. Thus, John G. Holden, one of Guy’s long-time collaborators, served as co-author on this response.

References
