More Than Meets the fMRI: The Unethical Apotheosis of Neuroimages

Eran Shifferman
Edmond J. Safra Center for Ethics, Tel Aviv University

Biography
Eran Shifferman is an independent researcher focusing on various aspects of the evolution of cognition. He began his academic journey as an ethologist and then switched to philosophy of biology and psychology, and finally ethics and science communication. In his work Shifferman attempts to provide broad palate accounts of the emergence, persistence, propagation and fixation of cognitive novelties along the phylogenetic tree. While doing so, Shifferman is critical of the many discourses he’s using and tries to make these scientific products more accessible to the public.

Acknowledgments
I’d like to deeply thank Snait Gissis and Aida Robles-Gomez for attentive and insightful readings; the fellows at Edmond J. Safra Center for Ethics in Tel Aviv University for their helpful comments; and Christine Schwabb and Gila Blitz for their comments

Publication Details

Citation
More Than Meets the fMRI: The Unethical Apotheosis of Neuroimages

Eran Shifferman

Abstract
The following is an attempt at a multifaceted critique of cognitive neuroscience’s use of images born of blood-oxygen-level dependent functional magnetic resonance imaging (BOLD-fMRI) in support of nesting cognitive functions in specific brain regions. It is an exploration of problems associated with three levels of producing functional neuroimages (NIs): the technological, the methodological, and the philosophical. My goal is not merely to map the spectrum of problems associated with the use of BOLD-fMRI NIs use in cognitive neuroscience. Rather, it is to use this map to support the claim that functional neuroimaging all too often amounts to unethical science, one where the generators of data overlook significant shortcomings of their tools of the trade and press forward with producing claims about the nature of the mind-brain link, which are too strong to be supported, by exploiting the strong appeal of their meticulously crafted images. These claims filter through to have a significant impact on the oblivious public over cardinal topics in psychology and philosophy such as behavior, emotions, consciousness, cognition, and the self.

Keywords
Neuroimaging, Ethics, Cognitive neuroscience, Reductionism, Evolution, Philosophy of Biology, fMRI, BOLD, Essentialism, Public Understanding of Science, Representation, Aesthetics

Introduction
Functional neuroimages (NIs) are vivid, colorful renditions of the brain based on raw numerical data generated by MRI scans taken as subjects perform a cognitive task, and their ensuing extensive statistical manipulations. Progressively, NIs have become a mainstay of neuroscientific arguments concerning the neural underpinnings of cognition and behavior, and have crossed over to the mass media as a canonical representation of the brain. Within the discipline of cognitive neuroscience, NIs are used to support the argument that (at best) particular networks or (more commonly) individual brain regions house specific cognitive, behavioral or emotional phenomena. The long list of cognitive and emotional traits that functional neuroimaging studies attempt to map onto the brain includes (but not limited to): belief (Harris, Sheth, and Cohen 2008; Harris et al. 2009; Kapogiannis et al. 2009; Neubauer 2014; Beauregard and Paquette 2006); humor (Bartolo et al. 2006; Chan et al. 2012; Sawahata et al. 2013); political orientation (Schreiber et al. 2013; Ahn et al. 2014); love (Bartels and Zeki 2000, 2004; Aron et al. 2005).
2005; Wan et al. 2014; Fusar-Poli and Broome 2007); moral behavior (Cikara et al. 2014; Yoder and Decety 2014); deception (Aharoni et al. 2013; Koster-Hale et al. 2013; Yang et al. 2014); happiness (Kong et al. 2015); “cultural” differences (Han and Ma 2014); and even stock market forecast (Smith et al. 2014) and response to reality TV shows (Melchers et al. 2015). In their review of NI studies, Gabrieli et al. (Gabrieli, Ghosh, and Whitfield-Gabrieli 2015) boldly claimed that the ability to predict future behavior using NIs is “a humanitarian and pragmatic contribution of human cognitive neuroscience to society”.

This paper brings together for the first time four disparate lines of criticism of various aspects of BOLD-fMRI research within cognitive neuroscience that dampen the stronger neurocognitive arguments concerning the neural basis of the mind. The technological criticism has two components to it: on the one hand is the physiological nature of the BOLD signal and its relation to neural activity, and on the other hand are the limitations and parameters of the MR machine itself. The second critique is methodological, dealing with the experimental designs and methods as well as statistical tools used to produce NIs, and the impact of the vast array of available design possibilities on them. The third line of investigation is the philosophical, addressing the basic tenets and widely accepted (and practiced) underlying assumptions concerning the link between neuronal activity - as portrayed by functional NIs - and claims about the nature of brain operation, the mind, and the self. Fourthly, the resultant criticism will then be merged with some sociotechnological insights concerning the concept of representation and the power dynamics between scientists and non-professionals. This is done in an attempt to complete and support the main argument of this paper: that the use of fMRI NIs in order to triangulate brain coordinates of consciousness and cognition is direly problematic and yet marketed to non-professionals as bona fide scientific truth that in turn shapes specious public perception.

The Physiological Basis of fMRI

To begin with, let us clarify what BOLD-fMRI allows for. Nerve cells’ activation is an energy consuming process that relies on the metabolism of glucose. These metabolic demands depend on blood flow and cause brain blood oxidation status changes. Oxygenated and deoxygenated blood have different magnetic properties, and to determine whether a brain region has been activated in response to a specific stimulus, one scans for an increase in oxygenated blood in that region as a function of conditions that change over time. Thus - crucially - the BOLD signal is an indirect indicator of neural
activity, in which blood and metabolism are the mediators. Cardinal to the validity of
the BOLD signal as an indicator of neural activity is neurovascular coupling, a metabolic
hypothesis of a one-to-one correspondence between hemodynamic changes (the BOLD
signal) and neuronal activity. However, evidence shows exceptions and variations to this
assumption, thus complicating the interpretation of fMRI data when analyzing cognition
and behavior (Caesar, Thomsen, and Lauritzen 2003; Devor et al. 2008; Sirotin and Das
2009; Ekstron 2010; Jukovskaya et al. 2011; Mishra et al. 2011; Hermes et al. 2012; Siero
et al. 2013; Huo, Smith, and Drew 2014; Mayhew et al. 2014). This has led some scientists
to warn against assigning too great an importance to fMRI reports (Page 2006; Rossier

**BOLD Problems**

The reliability of the BOLD signal as an indicator of neural activity is cast deeper
under shadow as there is more to brain metabolism than just neurovascular coupling.
First, firing neurons are not the only cells to cause hemodynamic changes: we can add
neurons at subthreshold levels of activation, neurons with varying levels of simultaneous
excitation and inhibition, and feedback from local and distant sites (Nair 2005). Non-
n neuronal entities such as astrocytes and vascular cells also impact brain metabolism
(Iadecola 2004; Bélanger, Allaman, and Magistretti 2011; Figley and Stroman 2011;
Escartin and Rouach 2013). Second, there is a broad range of different kinds of neurons,
each with different genetic makeup, neurotransmitter composition, myelination profiles,
spatial structure, connectivity topography, and spatiotemporal functionality, all shaping
neuronal operation (Fishell and Heintz 2013) and inflicting unique metabolic changes
(Logothetis 2008). Third, the BOLD signal primarily measures the input and processing of
neural information within a region and not the output signal transmitted to other brain
regions (Logothetis 2003). Without an output, a given brain region is highly unlikely
to be generating behavior. Moreover, input to an area, and processing within it, are
necessary for the disengagement of that region’s function, causing such disengagement
to also contribute to the BOLD signal (Page 2006). Fourth, BOLD contrast is generally
most intense in the veins downstream from the neural circuits that create metabolic
demand, causing the location of a signal change and the location of the presynaptic
neural activity to not necessarily correlate (Tancredi and Brodie 2007).

Taken together, the composition and characteristics of local neural activity impede
BOLD-fMRI’s ability to differentiate between function-specific processing and other
neural artifacts; between bottom-up and top-down signals; and between excitation
and inhibition. Also, the magnitude of the signal does not necessarily correlate with the importance of the respective region for the task of interest and cannot be standardized to quantify differences between brain regions, or between tasks within the same region (Huber 2009). It follows that a brain region, a neuroimage, and a cognitive function do not necessarily have a linear one-to-one correspondence. Rather, the relationship is best described as a statistical correlation (Page 2006; Tancredi and Brodie 2007; Logothetis 2008). Such a correlation does not suffice to justify a strong localization argument.

**BOLD-fMRI - Technological issues**

Exploring the nature of the BOLD signal allowed us to lay the foundation for discussing the first critique, that of the MR equipment itself. In this section we discuss technological hurdles put in front of cognitive neuroscientists when attempting to draw conclusions from BOLD-fMRI data to cognition and behavior1.

**Spatiotemporal Limitations**

There are two ways by which spatiotemporal dissonance between the actual neural activity and the ensuing BOLD signal may arise. The first is caused by apparatus limitations. A temporal limitation is levied due to the fact that there is a discord between the time required to generate a given cognitive function (several tens of milliseconds) and that required for an MR machine to collect enough raw data (2-6 seconds) (Haxby, Courtney, and Clark 1998). A spatial limitation is levied since MR has a resolution of about 0.1 mm, while single cell activity operates at three to four orders of magnitude smaller (Hardcastle and Stewart 2002)2. The second cause of dissonance is the fact that the BOLD signal is not a direct measure of neural activity, and thus the signal may be temporally and even spatially out of register with the activity changes that are ultimately the phenomenon of interest. In order to bridge this gap we have to make a variety of assumptions about the temporal and spatial relationships between blood flow changes and neural activity, the validity of which is at times questionable (Roskies 2008). Another

---

1. Logothetis’ (Logothetis 2008) excellent discussion of the plethora of shortcomings of current use of BOLD-fMRI in neurocognitive sciences should serve as reference for the many technical details this paper cannot fully explore.

2. New advancements allow for a significantly improved resolution and sample rate, yet these are not widespread and thus not enough to observe complex spatial-temporal orchestration of brain activity that underlies cognition (< 10 msec) (Buckner 2003). Therefore, what cortical processes exactly the BOLD signal does and does not represent is still far from clear (Goense, Whittingstall, and Logothetis 2012).
issue concerns brain anatomy: areas adjacent to sinuses (e.g. the orbital frontal cortexes) hold air inside, and this proximity may distort the image (Kringelbach and Rolls 2004), particularly with increasing magnetic field strength (Devlin et al. 2000).

**Resolution Problems**

A typical MRI voxel contains 5.5 million neurons, $\sim 4 \times 10^{10}$ synapses, 22 km of dendrites and 220 km of axons (Logothetis 2008). These astronomical numbers include a wide array of neuronal types, which necessarily forces heterogeneity upon any given region of interest. In addition, “flexible” neurons can represent different abstract rules or categories in a temporal and context-dependent manner (Duncan 2001), thus making the definition of the functional role of brain regions a cumbersome task. Furthermore, all the processing within a given voxel involves an extensive range of inputs from other brain regions (Hardcastle and Stewart 2002). Consequently, conclusions concerning the neuronal activity of a given voxel are governed by the haphazard content of a given sample, thus underestimating when neurons actually respond and under what conditions (Hardcastle and Stewart 2002; Roskies 2008; Meinertzhagen et al. 2009).

One way of tackling this problem is via improved resolution, which allows for more accurate reading of the hemodynamic metabolic changes. Accuracy is measured by the signal-to-noise ratio (higher ratios indicate better quality). The strength of the MR machine’s primary magnetic field is of paramount importance to the signal-to-noise ratio, such that stronger fields allow for greater acquisition of significant voxels (Hoenig, Kuhl, and Scheef 2005; Alvarez-Linera; García-Eulate et al. 2011; Wardlaw et al. 2011). Most MRI machines used in neurocognitive science operate at 1.5-3 Tesla (Logothetis 2008; Regatte 2014). However, not only are identified areas likely to morph and grow/shrink; some are likely to be statistically significantly active (if not maximally so) when investigated under higher magnet power, even if this activity is below threshold at 1.5 T (Figdor 2010). So it follows that at least some regions of activation established by using low-frequency scanners may actually be artifacts of their magnet power. Unfortunately, tinkering with any parameter, be it magnetic field strength or voxel size, also impacts negatively on the signal-to-noise ratio, and a delicate balancing game is required, all depending on the research question and study limitations. The reality is that there is a myriad of variables that have a direct and immediate impact on the final NI, and there is very little standardization that allows minimization between- (and even within-) lab procedural variability (Bennett and Miller 2010).
Shifferman

Admittedly, technological issues are being addressed\(^3\). However, the persistent problem, in the context of this paper, is the fact that the data already generated with older technology is not scrutinized and re-evaluated with improved technology (Vul and Pashler 2012), thus passively allowing possible erroneous data to solidify as a valid reference. Therefore, considering MR technology alone, we have ample reason to doubt strong localization arguments resting on such NIs (see the following discussion of replication, page 8).

**BOLD-fMRI - Methodological Impediments**

As stated earlier, neuroimaging concerns extend beyond technological facets. This section presents the most pertinent methodological tools used to spawn NIs and the difficulties associated with them: localization, raw data processing, statistical manipulation, variability, and standardization.

**Localization**

Localization is the attempt to nest a specific cognitive/emotional attribute in a specific brain region. Localization is a central goal of cognitive neuroscience and it is deeply rooted in the history of neuroscience, long before any mapping technology was available (Naneix 2009). Localization has waxed and waned, but the arrival of modern imaging techniques, primarily that of MRI, has brought it to new heights (Vul and Pashler 2012; Klein 2012). Importantly, with the exception of extensively studied peripheral cognitive areas, few neuroimaging localization attempts are beyond controversy (Figdor 2010; Ihnen et al. 2009; Ball et al. 2009). Here, I investigate the localization project by looking at its components and highlight major concerns associated with it.

**Subtraction**

Localization via BOLD-fMRI is made possible by subtraction. Subtraction entails imaging subjects performing a mixed sequence of two different tasks that are (supposedly) separated by a single cognitive element, ending up with two different time series that can be compared to verify whether the activity in the region of interest was different between the two tasks. Once performed, the image of the “simpler” task is subtracted from the more complex one, creating a difference image that (ideally) has

---

3. E.g., cutting edge fMRI technology today allows for a single-cell resolution with 17.2T (Radecki et al. 2014). Currently, even 7T is not FDA-approved, though it’s popularity rapidly increases (Kraff et al. 2015).
isolated an area of increased or decreased activation. That area is considered to be the seat of the additional cognitive element separating the two tasks (Vul et al. 2009). Subtraction relies on the pure insertion assumption: the amount of additional activity attributable to the interaction between the new and the old tasks is zero. Subtraction hinges on several strong tenets: a) cognitive processing is highly modular; b) the brain is a serial processor; c) cognitive functions are linearly additive so there are no qualitative changes upstream on the shared components of experimental and control tasks and can therefore be subtracted from one another; d) each task invokes a minimum set of components for successful performance (van Orden and Paap 1997; Peterson 2003). At the neural level, the assumption is that the difference in neural activity during baseline and task is due entirely to the new task and does not represent any influence on or interaction with the baseline activity. This assumption ignores the possibility that the additional neural activity may be neither sufficient nor necessary for the presumed-to-be purely additional task (Figdor 2010).

Furthermore, as demonstrated earlier, due to the nature of the BOLD signal and the technological limitations of MR sampling, subtraction cannot determine whether the differences in activity are due to the cognitive process assigned to them a posteriori or due to something else occurring concurrently but coincidentally. Ironically, the lower the sensitivity of the MR apparatus, the better it is for localization: low signal-to-noise ratios allow for only a few statistically significant differences across conditions to be found (Hardcastle and Stewart 2002). Also, pure insertion does not fit well with observations indicating that neural processing utilizes spatiotemporal feedback connections between multiple regions (Poldrack 2010), including even partial information transfer between stages of processing (Miller and Hackley 1992; Bichot, Rao, and Schall 2001). If the pure insertion assumption fails, then there is no way to determine what cognitive processes are reflected in the activation observed in the subtraction experiment.4

Reverse Inference

Even if cognitive neuroscientists were to dispense with the problems associated with pure insertion, they would still have to address seriously another major concern of localization, that of reverse inference: inferring the operation of a specific cognitive trait

---

4. These criticisms have not gone unnoticed, and cognitive neuroscientists have invested effort in adopting designs that attempt to circumvent pure insertion (Poldrack 2010; Kawabata Duncan and Devlin 2011). However, Poldrack (Poldrack 2010) argues that the new methods fail to establish a link between task characteristics and the cognitive processes they are supposed to represent.
based on activation of a specific brain region, reasoning backwards from activation to cognitive operation. For example, activation in the amygdala is interpreted as reflecting fear or negative emotion, even though it can be equally active for positive outcomes (Bunzl, Hanson, and Poldrack 2010). Several observations cast a shadow over reverse inference. First, pluripotency\(^5\) virtually negates reverse inference because activity of a region in response to two or more contexts cannot be used as evidence assigning that region exclusively to only one of them (Haxby et al. 2001; Henson 2005; Price and Friston 2005). Second, evidence shows that the response of regions other than those responding maximally to a given stimulus could also predict which stimulus was presented (Haxby et al. 2001; Mole et al. 2007). Together, this evidence shows that reverse inference reflects the logical fallacy of affirming the consequents (Poldrack 2006; Klein 2012).

**Additional Concerns over Localization:**

Going beyond the problematic pure insertion and reverse inference, other shortcomings of localization have been highlighted. First, traditionally, fMRI studies perform only a handful of scans of many subjects over relatively short time period. Contrary to that, Gonzalez-Castillo et al. (Gonzalez-Castillo et al. 2012) scanned very few subjects for a long cumulative duration, over many scanning sessions. Their analysis shows that 70%-90% of all voxels were labeled as active, suggesting that localization is made possible due to insufficient statistical power. If, as Gonzalez-Castillo and colleagues (Gonzalez-Castillo et al. 2012) suggest, the entire brain is involved in even minor tasks, then the dichotomy between active and inactive regions is of no scientific relevance (Stelzer et al. 2014), and the localization project is severely weakened.

Secondly, early meta-analyses have shown that localization is elusive since the execution of cognitive functions relies on networks connecting many regions with a high degree of spatiotemporal variability, inter-individual differences and context-dependence (McIntosh 2000; Cabeza and Nyberg 2000; Phan et al. 2002; Gerlach 2007; Buchsbaum and D’Esposito 2008; Dolcos, Denkova, and Dolcos 2013).

Cumulatively, the idea of a one-to-one mapping of cortical activation to high-level cognitive processes suggested by NIs seems like an oversimplification of a more complex

---

\(^5\) Pluripotency is the ability of the same neural entity to perform (or cause or be involved in the performance of) several tasks (Nair 2005; Price and Friston 2005; Henny et al. 2012; Lee, Soares, and Beique 2012). Pluripotency connects with the philosophical concept of multiple realizability (the claim that the same mental attribute can be generated by more than one physical substrate) and to the biological concept of degeneracy (the ability of structurally different elements to yield the same output; (Edelman and Gally 2001)).
many-to-many mapping (Just and Varma 2007). This is exemplified by a letter sent to *The New York Times*, in which a group of nearly 20 prominent cognitive neuroscientists wrote:

We know that it is not possible to definitively determine whether a person is anxious or feeling connected simply by looking at activity in a particular brain region. This is so because brain regions are typically engaged by many mental states, and thus a one-to-one mapping between a brain region and a mental state is not possible. (Lavazza and De Caro 2010)

This brings us to another crucial problem of localization: if brain qualia such as pain are to be reduced and decomposed to neural process, they must first be given a functional definition or else the reduction enterprise fails (Harley 2004; Kim 2006). Such a definition can only stem from a robust theoretical basis. However, there is no theoretical (psychological or philosophical) foundation for localization claims (Uttal 2002; Gerlach 2007; Poldrack 2010; de Graaf, Hsieh, and Sack 2012; Rathkopf 2013). The ability to infer about neural correlates of a cognitive process is often confined by the conceptualization of the process into a task that can be performed in the scanner (Bell and Racine 2009). Neuroimagers, while refraining from addressing the necessary psychological theoretical discussion, still, through their experimental design, reflect their own interpretation of specific cognitive functions (Burock 2009; Huber 2009; Roskies 2010). Many worry that such interpretations of functional decompositions demonstrate a naïve understanding of the cognitive processes underlying the performance of complex tasks (van Orden and Paap 1997; de Graaf, Hsieh, and Sack 2012; Aru et al. 2012). Some go as far as claiming that neurological data is irrelevant to cognitive psychology until a complete psychological theory has been established, at which point the neuroscientific data would be redundant (Harley 2004; Coltheart 2006, 2004; Loosemore and Harley 2010).

It is important to remember that the malleability of cognitive neuroscience theories is a direct outcome of the pliable nature of several psychological and psychiatric theories. Both disciplines have allowed some research within them to lax its scientific rigor to a degree where some warn of an intellectual crisis (Fava 2006). If the psychological theories guiding the localization project are themselves to be doubted, what stock can be afforded to the theories cognitive neuroscience develops using functional NIs? (John,

---

6. Psychology’s proclivity for severely exaggerated high rate of positive results is well documented (Fanelli 2010; Francis, Tanzman, and Matthews 2014).
Shifferman

Loewenstein, and Prelec 2012; Yong 2012). By using a concrete example (Kanwisher et al.’s localization of facial recognition to the brain region known as FFA), Mole et al. (Mole et al. 2007) claim that such studies are simply not new or illuminating: “It has told us that if there are special resources for the processing of faces then the FFA is a likely site for them. But scanning can do nothing to answer the question of whether there are such resources”.

Localization is a particular case of reductionism, and therefore must be contextualized as such. In the attempt to analyze complex concepts using NIs, a dramatic simplification and reduction of the study objectives must occur in order to attain a feasible experimental protocol (Huber and Huber 2009). This deflation is interesting not solely in terms of the scientific process, but also because it may pertain directly to a possible human cognitive feature that prefers simpler accounts, which could explain why reductionism appeals to us more strongly than holistic or complex accounts (Rose 1999; Bunzl, Hanson, and Poldrack 2010).

In lieu of the above, it would seem that localization is more of an experimental choice than a scientific model, a way of making sense of this entangled mass called the brain. There appears to be a tacit agreement within some circles of the cognitive neuroimaging community that the claims hatched within it are best described as heuristic placeholders. Such voices acknowledge that while localization is limiting, it is still an important step that, once corrected, can lead to better understanding (Mundale 2002; Bechtel 2002, 2004; Craver 2005). If localization is merely such a tool (one with experimental advantages but not necessarily a reflection of actual brain activity), then it brings to the surface the tension between knowledge shared within the community and that disseminated outside of it. This begs the ethical question that if a discipline knows that it employs tools for heuristic purposes, then why does it insist on knowingly generating public claims too strong to be supported?

**Standardization**

Another problematic methodological tool employed in neuroimaging is a consequence of the biological reality that no two brains are alike, neither anatomically or functionally (Miller et al. 2012). Despite more than a century of research, there is still no consensus on reliable delineation of functional subdivisions in the brain, mostly

---

7. Reductionism has generated a by now insipid platitude of interpretations far beyond the scope of this paper. Readers can enrich their knowledge of reductionism by referring to the droves of publications discussing it.
because there are no binary abrupt transitions between brain regions (Peterson 2003; Haueis 2012; Cox et al. 2014). This has brought about the practice of standardization: the process of either averaging out the scans of a given study or comparing all subjects from the same experiment to an established standard atlas. Both forms of standardization are precarious since an averaged brain represents all brains and none, akin to Quetelet’s *l’homme moyen*. In addition, researchers rarely provide information as to the way the norm was created (i.e. the characteristics of the subjects used to that end; (Reeves et al. 2003)). Furthermore, some of the most popular atlases used for standardization are not only many decades old but also based on the anatomy of a single subject (Bogen 2002). Thus, the epistemic value of NIs depends upon whether the idealized brain it portrays is representative of real brains with regard to the anatomical, physiological, and psychological factors relevant to the question under investigation (Bogen 2002; Uttal 2013).

Going beyond its legitimacy, standardization distorts structure/function analyses: in the processing phase of creating a NI, brain regions with different functional profiles near the region of interest are averaged together across individuals, reducing both the resolution and the sensitivity of subsequent functional analyses (Saxe, Brett, and Kanwisher 2010). Thus, function cannot be assigned purely on the basis of spatial patterns (Sadaghiani et al. 2010). Unfortunately, this is too often left out of neuroimaging discussions (Jbabdi, Sotiropoulos, and Behrens 2013), and is utterly absent in mass media reports.

### Processing

For cognitive neuroscience to generate general observations concerning cognition and behavior it must compute correlations across subjects (Roskies 2008), and then each individual brain scanned has to be mapped onto an average brain. For this, the raw time series must undergo preprocessing to reduce noise. Vul & Kanwisher (Vul and Kanwisher 2010) describe the highly complex process required to convert raw data into publishable NIs:

The time series of voxel changes may be motion-corrected, coregistered, transformed to match a prototypical brain, resampled, detrended, normalized, smoothed, trimmed (temporally or spatially), or any

---

8. Once we throw in pluripotency, we critically restrict our ability to distinguish what regions do as a whole and what sub-regions do individually (Bogen 2002).
subset of these, with only a few constraints on the order in which these are done. Furthermore, each of these steps can be done in a number of ways, each with many free parameters that experimenters set, often arbitrarily. After preprocessing, the main analysis begins. In a standard analysis sequence, experimenters define temporal regressors based on one or more aspects of the experiment sequence, choose a hemodynamic response function, and compute the regression parameters that connect the BOLD signal to these regressors in each voxel. This is a whole-brain analysis, and it is usually subjected to one of a number of methods to correct for multiple comparisons… the whole-brain analysis is often the first step in defining a region of interest in which the analyses may include exploration of time courses, voxelwise correlations, classification using support vector machines or other machine learning methods, across-subject correlations, and so on. Any one of these analyses requires making crucial decisions that determine the soundness of the conclusions.

This detailed description shows that BOLD-fMRI NIs represent mathematical constructs rather than physiological reality (Burock 2009). The abundance of mathematical processing applied to the raw data leads to a skewed representation and estimation of many neural activities directly pertaining to the cognitive processing of a given task. Thus, the nature of processing alone demands great caution in interpreting functional NIs in cognitive neuroscience context (Logothetis et al. 2001). Consider spatial smoothing as an example: after smoothing, each voxel contains a mix of its own signal and the weighted signal of surrounding voxels. The justification for averaging the BOLD signal over space is improving statistical sensitivity. At the same time, spatial smoothing generates a systematic bias of spatial localization (Sacchet and Knutson 2013), as separate and distinct activations progressively blend into one another (Geissler et al. 2005). Stelzer et al. (Stelzer et al. 2014) argue that more than 90% of the post-smoothing signal at any given location originates from neighboring voxels, thus increasing the numbers of false positive voxels. These authors went as far as stating that due to spatial smoothing it is impossible link fMRI data with data from other neuroscience disciplines.
How Statistical Tools are Used

The convoluted process of generating a functional NI does not stop with mathematically transforming raw data via processing. Since raw data are an astronomical amount of numerical values in long time-series, it is imperative to perform statistical operations in order to convert them into images. However, statistics, paraphrasing D’Israeli, are the most mendacious of all lies, and since the choice of statistical tools has a direct and paramount impact on the resultant image and the conclusions that can be drawn from it. And the list of available statistical tools in cognitive neuroscience is impressively long (Carp 2012a, 2012b).

Before we delve into the role statistics play in BOLD-fMRI, we have to present two basic definitions. A type I error falsely rejects a true null hypothesis and generates a false positive: accepting that a hypothesized event exists when it does not (e.g. a wrong medical diagnosis). A type II error is the acceptance of a false null hypothesis, yielding a false negative. Importantly, false negatives are correctable with ensuing research, whereas false positives are difficult to refute once established in the literature and not re-evaluated properly (Bennett, Wolford, and Miller 2009). The various factors contributing to elevated false positive rates in BOLD-fMRI (Bennett, Wolford, and Miller 2009) can be brought under the umbrella of poor application yielding low statistical power, which in turn complicates replication and fosters contradicting conclusions in the analyses of the same database (Duncan 2001; Button et al. 2013; David et al. 2013).

A ubiquitous statistical error in functional neuroimaging is the non-independence error (aka double dipping): using the same data for selecting the voxels of interest and then using these voxels for the secondary analysis, the one upon which the functional conclusions are based9. Double dipping violates random sampling because the test statistics are not inherently independent of the selection criteria of the region of interest, thus statistically guaranteeing the outcome of the second analysis and rendering them useless (Kriegeskorte et al. 2009; Vul et al. 2009). Similarly, as mentioned before, statistical tests in neighboring voxels are not independent of one another, because time series in neighboring voxels are intercorrelated (Peterson 2003). Analyses have shown that the non-independence error is widespread in BOLD-fMRI studies (40-50% of published papers) and that the severity of the distortions of the results presented in these papers could not be assessed. This necessitates replications and reanalysis (Kriegeskorte et al. 2009) or the results of these studies “mean almost nothing”, since they are “using

9. It must be noted that this problem is not unique to cognitive neuroscience, as it is widespread in neuroscience and psychology too (Fiedler 2011).
seriously defective research methods and producing a profusion of numbers that should not be believed” (Vul et al. 2009).

In addition to double dipping, there are other ill-used statistical tools in cognitive neuroscience. A major such source of complication is the necessity to correct for multiple comparisons: many researchers find these corrections too draconian, and choose to either avoid correction altogether, or to employ lenient statistical tools (Saxe, Brett, and Kanwisher 2010). Bennett et al. (Bennett et al. 2011) used an extreme test case to demonstrate that by using uncorrected statistics for multiple comparisons they were able to generate a NI that showed active voxel clusters in the brain of a dead fish in response to visual stimuli. The authors concluded that it is likely that “investigators do not want to jeopardize their results through a reduction in statistical power”.

Another case in point is the statistical dichotomy between significant and non-significant results based on P values. This comparison often erroneously involves two separate tests in which researchers conclude that effects differ when one effect is significant (P<0.05) but the other is not (P>0.05), while the comparison should be between them (Nieuwenhuis, Forstmann, and Wagenmakers 2011). Numerous (and early) articles have clearly demonstrated that this dichotomy is arbitrary and unwarranted, as any strong evidence against a null hypothesis (if such at all exists) depends on other conditions and cannot be expected to be globally valid at p<0.05 (Sterne, Cox, and Smith 2001; Wacholder et al. 2004; Ioannidis 2005b). Nonetheless, this did not prevent the spread of the statistical error common in neuroimaging studies of comparing significance levels (Henson 2005; Poldrack et al. 2008). Nieuwenhuis et al. (Nieuwenhuis, Forstmann, and Wagenmakers 2011) found that this dichotomy is prevalent even in high profile journals, and that in some cases the error may have contributed substantially to the article’s main conclusions.

Yet another sizeable statistical concern is unfitting sample sizes: most published fMRI studies have sample sizes that would be considered exceedingly small by conventional standards (Yarkoni 2009; Button et al. 2013; Ingre 2013), if they include sample size calculations at all (Guo et al. 2014). It is established that in fMRI studies, small studies (n=16) fail to reliably distinguish small and medium-large effect sizes from random noise as do larger studies (n=100) (Ingre 2013)10. However, Wager et al. (Wager et al. 2009) report that across 415 fMRI studies reviewed, the average group size was smaller than 12, with some using only 4 subjects. At the same time, the number of activation loci claimed to be discovered by them is relatively large (David et al. 2013).

10. Zandbelt et al. (Zandbelt et al. 2008) provide a sample size estimations for BOLD-fMRI crossover studies.
This statistical bungle is exacerbated by the observations that most labs employ statistical tools according to historical precedent rather than through formal power calculation (Button et al. 2013). These statistical methods were developed to allow fMRI to detect activation rather than characterize it, thus making the interpretation of results often speculative (Monti 2011). Unfortunately, most fMRI researchers have only a vague idea of how reliable their results are, and the more tasking cognitive attributes are the ones with the lowest fMRI reliability (Nichols and Hayasaka 2003; Bennett and Miller 2010; Saxe, Brett, and Kanwisher 2010). In fact, Uttal (Uttal 2013) argued that “many statisticians would be amused by the cavalier attitude of some neuroscientists in assuming that their data meet the most basic criteria for statistical robustness” (p. 55).

Ioannidis (Ioannidis 2005b) lists six parameters diminishing the probability that statistical findings in functional NI are valid: 1) small studies; 2) small effect size; 3) the greater the number and the lesser the selection of tested relationships; 4) high design and analysis flexibility; 5) financial stakes and other biasing elements; 6) a hot field drawing many labs to it. Cognitive neuroscience falls short on all these criteria, thus casting a looming shadow over their produced claims.

Another crucial example of lenient scientific austerity, a direct amalgamated result of the methodological shortcomings listed above, is the alarmingly low rate of experimental replication in both psychology and fMRI studies (e.g. (Pashler and Harris 2012)). Moreover, when replication does take place it often contradicts initial reports, particularly if those were based on small samples size and/or published in high impact factor journals (Ioannidis 2005a). As shown earlier, conclusions drawn in studies committing methodological and statistical errors can continue to propagate and serve as basis for future null hypotheses because older studies are rarely re-evaluated and the publication process is biased toward positive results. The accumulative effect of many such distortions, regardless of their magnitude, is a grave impact on the validity and robustness of localization claims.

The wide range of tools and analyses that can be operated on the full arsenal of methods applied in BOLD-fMRI research leads to a widespread phenomenon throughout science, that of high analytic flexibility and selective analysis reporting: choosing the most favorable experimental/analytical combination, the one that promote positives results (Carp 2012a; Button et al. 2013).
Variability

The final methodological concern is results variability, which runs the gamut from within- to between-subjects and between labs. The obstacle is that such variability is yet another variable hampering the replication and attainment of consistent results (Uttal 2013). Early fMRI studies have documented intra-subject variability, even after repeated tests in the same laboratory over a number of days (Zandbelt et al. 2008), particularly with cognitive tasks (McGonigle et al. 2000). Later work has unveiled inter-subject variability, and showed it to be greater than the intra-subject one (Miller and Van Horn 2007; Miller et al. 2009; Diederen et al. 2013; Tancredi and Brodie 2007). Additional variability exists between different laboratories: while individual experiments identify only a relatively small number of activation peaks per cognitive task, collecting all responses across many centers tackling the same cognitive attribute generates a distribution map covering the entire brain. Inter-venue variability is apparently so great that meta-analyses only exacerbate the situation and increase variability (Uttal 2013; Fox et al. 2015). The cumulative effect of these types of data variability is a serious impediment on the localization project, suggesting that there are no macroscopic-level delineations corresponding to cognitive performance, and that they are probably a methodological artifact (Gonzalez-Castillo et al. 2012; Thyreau et al. 2012).

Summing up, this overview of central and ubiquitous BOLD-fMRI methodologies demonstrates that the experimental design of functional neuroimaging studies (in addition to the restrictions imposed by the technology itself) acutely delimitates strong localization claims for pinpointing the neural substrate of cognitive functions. While some technological and methodological advancements have presented themselves throughout the years, they had alleviated mostly minor concerns. Methodologically speaking, the experimental rationale has remained mostly intact, and the philosophical concerns - which constitute the very keel of cognitive neuroscience’s arguments - still linger on and represent inherent flaws looming large over the validity of the localization project. Nonetheless, the allure of fMRI has attracted many scientists from different disciplines to use it in their work, and too many of them prematurely capitalize on established protocols rather than addressing their particular scientific needs (Pan et al. 2011). Thus, a growing number of neuroimagers are nescient with respect to the complexities and problems associated with BOLD-fMRI and the inner workings of MR machines and their capabilities. At the same time, the physicists and mathematicians responsible for improving MRI technology lack an intimate knowledge of cognitive and neurological
theories (Peterson 2003; Seixas and Ayres Basto 2008). This yields a dialogue of the deaf between producers and users, a situation not conducive to proper scientific practice.

**BOLD-fMRI - Philosophical Issues**

As stated in the opening of this paper, the problems associated with the BOLD-fMRI NIs for the purposes of cognitive neurosciences stem from a quadrumvirate of levels, each posing difficulties that are hard to alleviate. We have seen the kind of difficulties associated with the BOLD signal itself and the technological parameters of the MR machinery. Then we discussed experimental design and statistical manipulations and learned that they are significantly harder to allay. In this next tier, I focus on concerns emanating from philosophy of mind and philosophy of biology that constitute the most tenacious opposition to the localization project.

**The Mereological Problem: Psychophysics Revisited**

Neuroscientists are usually materialists that vehemently deny there is more to the mind than what the brain has to offer. The hubris of neuroscientists reflecting utter confidence in their ability to solve all things mind is best exemplified by Francis Crick’s statement “No longer need one spend time attempting… to endure the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem” (Crick 1996).

When neuroscientists claim to have discovered the neural correlate of a cognitive trait, the fundamental question from philosophy is what is it really that they show. For such a claim to be adequate, an isomorphism between neuronal form and function and experiential content, at a not-established description level, must exist (Noë and Thompson 2004). As demonstrated earlier, such a one-to-one correspondence, at least via functional neuroimaging, is currently not even technologically feasible. Furthermore, philosophers seriously doubt the validity of such future argument, even if better technologies were to present themselves (Sprevak 2011). Their counter argument is that feelings are felt, experiences experienced, thoughts created, and behavior displayed only at the level of the whole person interacting with her environment (Noë and Thompson 2004; Burock 2009). It is the person that cognizes, not her sub-personal organs, tissues, cells, organells, or molecules; not even if they are called the brain, the cortex, the amygdala, neurons, synapses, glial cells or dopamine. This is the mereological fallacy: assigning function to a part of a whole that is attributable only to the whole itself (Bennett and Hacker 2003; Pardo and Patterson 2010).
The psychophysical and mereological problems boil down to the vernacular question “are we our brains?”. As we have seen in the discussion of localization, reductionism has been the scientific bon ton for many decades now. Neuroscience offers some aggressive forms of reducing the mind to the brain, such as equating mental processes with neural processes or arguing that mental processes are causally inert epiphenomena of neural processes (Beauregard 2009). Fortunately, not all philosophers or neuroscientists subscribe to these points of view. Some philosophers reject reductive materialism by arguing that the brain cannot participate in the sensory and in the social (Burwood 2009), and that mental processes exert a causal influence on the brain (Paquette et al. 2003; Beauregard 2009). Recent scientific work strongly suggests that the body affects the development, homeostasis, and plasticity of the nervous system (e.g. (Qureshi and Mehler 2013)). Such findings have led some to argue that neuroscience is, ironically enough, a dualist enterprise: while rejecting a dichotomy between brain and mind, they ignore somatic effects on the brain, thereby effectively creating a dichotomy between the brain and the body (Glannon 2009).

If the rebuttal of a mandatory body-brain-mind-Umwelt complex is valid (Chiel and Beer 1997; Byrge, Sporns, and Smith 2014), then not a single component of this complex is sufficient by itself to generate and explain cognition and consciousness, as each carries only a proportional weight within that complex (Glannon 2009; Pardo and Patterson 2010; de Graaf, Hsieh, and Sack 2012). A useful analogy comes from another field that was dominated by fierce reductionism: genetics. Maybe if we think in terms of genotype (the brain, the neurotype?) and phenotype (the mind, the cognitype?) we could better explain why neuronal operations are necessary but not sufficient to explain the mind. As such, there is no 1:1 correspondence between neurotype and cognitype, as the neurotype serves only as a scaffold upon which the cognitype builds and elaborates via dynamic reciprocal interactions with the body and the environment. If we combine this with the persistent criticism expressed against cognitive neuroscience’s lack of psychological and cognitive theoretical background ((Uttal 2002); see page 6) we end up with Coltheart’s (Coltheart 2004) statement: “No amount of knowledge about the hardware of a computer will tell you anything serious about the nature of the software that the computer runs. In the same way, no facts about the activity of the brain could be used to confirm or refute some information-processing model of cognition” (p. 22).
Form and Function

The study of the dynamic relationship between form and function and the constraints they levy on each other has fascinated biology for centuries (Mundale 2002; Wouters 2005). Neuroscience is no exception and contributes its own questions to this debate (e.g. (Meinertzhagen et al. 2009; Friston et al. 2010)). Cognitive neuroscience presents a philosophical challenge because it demands that we explicitly define the form and function of consciousness in order to be able to properly design an experiment that will accurately identify its neural loci. Such definitions break down to questions such as what is a brain function in general or what is the function of a given brain area. Lamentably, this brings us back to the severe lack of theory in functional neuroimaging research. The form-function relationship is probably the crux of the difficulty functional neuroimaging has in establishing its claims: as long as there is no full mapping of what functions are served by which brain region and as long as the boundaries of these regions cannot be precisely delineated, no trustworthy localization claims can be made. Things get murkier when we remember that NIs address multiple neuronal levels of organization, each characterized by different expressions of form and function. Therefore, for functional neuroimaging to have a legitimate seat in the mind/brain debate, it has to meet with two prerequisites. First, at the very minimum, there has to be a clear definition of what brain/cognitive functions are for each and every experiment. Second, an explicitly detailed account of the neural substrate, be it localized or distributed, of that function must be given. However, this harks back to the problem that by answering these questions, any additional neurocognitive data would be redundant (see page 5 and 11).

After touching upon some philosophical concerns related to the neuroimaging practice of localization and its use in establishing claims about the nature of the brain-mind link, glaring deficiencies in the theoretical foundation of this scientific pursuit appear. Unlike the technological and methodological aspects discussed earlier, the philosophical lacunae are very difficult to alleviate, especially without overhauling the entire discipline. Therefore, it seems unlikely that a major revision of the core tenets of NIs’ use in the study of cognition will present itself in the near future. A corollary of this observation is that the scientifically dubious knowledge gained so far from this discipline will continue to proliferate unbridled from the corridors of academe through to office buildings coolers.

As an interim conclusion, I describe functional NIs’ inadequacy to support localization by alluding to the degrees of separation a NI has from the biological phenomenon it allegedly represents. The first degree of separation stems from the nature of the BOLD signal itself: the tremulous nature of the neurovascular coupling hypothesis prevents a
definite access to the neural activity measured. A second degree is introduced owing to the technological parameters of the MR scan, which force a spatiotemporal dissonance with the neural activity measured. Methodological choices force the next degrees of separation: the third separation results from violation of the pure insertion assumption, which then prevents us from linking a specific cognitive attribute to a brain region. A fourth degree enters the equation via methodological shortcomings, ranging from the failure of reverse inference, through the lavish baggage of processing, statistical manipulations, standardization and culminating with all the processes designed to generate a smoother and cleaner image by discarding or hiding data that does not fit well with the a posteriori assignment of the region-cognitive function link promoted in a given study. These degree force us to sincerely doubt the validity of the localization endeavor. Finally, the fifth degree results from the aesthetic predilection of neuroimagers (see p. 13). Thus, owing to these degrees of separation, NIs presented as a depiction of cognition and consciousness actually have only a gossamer tenure with neuronal reality.

Ethical Considerations

The ultimate goal of this paper is not merely to map the spectrum of problems associated with the use of BOLD-fMRI NIs use in cognitive neuroscience. Rather, it is to use this map in support of the argument that this practice all too often amounts to unethical science, one where the generators of data overlook known shortcomings of their tools of the trade and press forward with producing claims too strong to be supported by exploiting the strong appeal of their meticulously crafted images. These claims filter through and find their way to non-professionals and policy makers who are oblivious to the tangled web of complexities surrounding these captivating images and who lack the tools to doubt the conclusions attached to them.

Having discussed the scientific frailty of using NIs for studying consciousness and behavior, in this section I focus on two aspects of neuroimaging practice bearing ethical impact. First, I establish and explore the implications of the absence of a theoretical framework linking neuroanatomy with all things mind and self. Second, I investigate the role representation plays in the perception of functional NIs and the propagation of the messages they attempt to convey.

Lack of Theory

Ample examples were given throughout this paper (e.g. localization, p. 6) and are given in this section as to the understanding, covering both proponents and opponents
of neuroimaging’s use in cognitive neuroscience, that the field lacks a theoretical background to lean on. This refers to the absence of both a consensual psychological framework of cognition (van Orden and Paap 1997; Uttal 2002; Harley 2004; Coltheart 2006; Poldrack 2010; de Graaf, Hsieh, and Sack 2012; Klein 2012; Rathkopf 2013; Reiner 2011; Fox and Friston 2012) as well as a well-established unified theory of how the brain works (e.g. (Haxby 2010; Power et al. 2010)). Let us begin with a fundamental conundrum, a critique of physicalism (the identification of mental states with brain states): if a creature without a brain can think, thinking cannot be a brain state (Block 1996). To answer this intellectual exercise we must define what a brain is and what constitutes a thought. There are vast and profound differences between the mammalian, piscine, and insect brain, and still some insect and avian species outperform some mammal species in various cognitive tasks (e.g. (Shifferman 2011)). So, what is a brain? Is it the highly distributed and restricted nervous system of the ant or bee? Would the nervous system of cephalopods qualify? What is the neuronal communality that allows profoundly different nervous systems to still generate the (seemingly) same behavior, and how could it be such simple such systems outperform advanced systems in particular tasks?

This has led Uttal (Uttal 2002) to argue that it is impossible to define the cognitive attributes to be localized without circularity and imprecision, which, in turn, inevitably lead to erroneous localizationist claims. It also brought V.S. Ramachandran, a prominent psychologist, to opine that “98% of brain imaging is just blindly groping in the dark” (Dingfelder 2008). The lack of theory is further aggravated by the fact that we are constantly learning new things about very fundamental aspects of brain function11. A prime example of the combination of both the lack of theory and the constantly developing body of knowledge is the case of the body-brain link discussed earlier (p. 10).

Another bias in neuroscientific research that weighs heavily on the validity of localization conclusions is the focus on event-related activity, knowingly overlooking spontaneous neuronal activity, the intrinsically generated brain activity that is not attributable to specific stimulus (Fox and Raichle 2007). The average adult human brain consumes 20% of all the energy consumed by the body, yet event-related induced energy consumption accounts to less than 5% of the baseline level of activity (Raichle and Mintun 2006). Thus, to understand the brain we must overcome this metabolic bias and consider the component that consumes most of the brain’s energy: spontaneous neuronal activity.

11. A very recent random yet significant example is that different neurons have different profiles of longitudinal myelin distribution, thus directly shaping its communication range and abilities (Tomassy et al. 2014).
activity (Sadaghiani et al. 2010; Fox and Raichle 2007). Spontaneous brain activity fluctuates within and between different modes and should not be considered noise, as it is coherently expressed in larger neuronal populations and functionally meaningful (Laufs et al. 2003). Perhaps the most studied example is the default mode network, an anatomical assembly of multiple brain regions supporting the “stand by” state of alertness (Raichle et al. 2001). It is argued that this base level is designed to maintain and support a dynamic shift between an introspective, self-referential mode of mental activity to an extrospective, preparedness mode that remains alert to environmental changes (Fransson 2005). Thus, rest is a state of a continuous orchestrated activity that is intermittently overridden once a goal-oriented activity emerges in response to certain stimuli (Fransson 2005), which persists through active cognitive task performance (Fox and Raichle 2007). Hence, before we bridge this knowledge gap and understand better spontaneous neuronal activity, no localization argument of substance can be made.

Augmenting the deep problem of lack of theory in the design and analysis of functional NIs is the fact that other disciplines within neuroscience present alternative and opposing interpretations of the neural basis of cognition. One can imagine here a pincer movement, wherein the cognitive neuroscience narrative of cognitive functions sequestered to specific brain regions is challenged simultaneously but differently from both top and bottom. From the lower organization level perspective, some shift the reductive fulcrum to single cells (individual neurons or groupings of identical neurons) and argue that they alone suffice to support the execution of some cognitive functions [e.g. (Smith and Ratcliff 2004; Nieder and Merten 2007)]\textsuperscript{12}. The attack from north points to the brain as a highly interconnected, spatiotemporal dynamic system that uses distributed representational schemes and relies on contextual and often transient sharing of neural resources across tasks in asynchronous and parallel fashion (Fingelkurts, Fingelkurts, and Kähkönen 2005; Fox et al. 2005; Henson 2005; Nair 2005; Sporns 2014; Zeki 2015). Proponents of networks as the foundation of cognition argue that there is extremely limited evidence supporting local non-linear neuronal operations, and these examples can be explained alternatively by using higher-level neuronal elements and emergence (e.g. (Bermúdez i Badia, Bernardet, and Verschure 2010)). They further assert

\textsuperscript{12} A much studied and criticized example of single cell-based cognition is grandmother cell theory. These are neurons argued to be solely responsible for the neuronal response for a stimulus, as they respond only to a very specific stimulus using neural convergence, in which neurons compute their various inputs in to a complex representation of a specific percept (see (Gross 2002; Quian Quiroga et al. 2008; Quian Quiroga et al. 2005)).
that the composite downstream effect of these neuronal assemblies (i.e. cognition), cannot be achieved either by single neurons alone (Buzsáki 2010) or by a brain region (Petersen and Fiez 1993).

Clearly, these accounts are mutually exclusive, thus reiterating the dire need for a comprehensive neural theory consolidating the abundant neuroscientific and psychological data and models. Also evident from glimpsing other disciplines of neuroscience is that a single neuronal element cannot, by itself, be a sufficient explanation either for the operation of neuronal systems or for the cognition/behavior it supposedly supports. If that were the case then the function of the entire nervous system would have collapsed into the operation of that single element, thus nullifying the need for a system. In that neuroecological context, functional NIs fail as they ignore both lower and higher organization levels: at the nerve cells level it is technically blind to a rich arsenal of microcircuitry, while at the regional level it overlooks the fact that there is no simple correspondence between structural and functional domain boundaries (Damoiseaux and Greicius 2009). This state of affairs has led Reiner (Reiner 2011) to assert that

Until we have a satisfying mechanistic account of how two similar neurons become distinguishably specialized in their selectivity it is best if arguments concerning how brains work adopt a dash of reservedness, one that realizes the limitations of our technology and method and acknowledges that all we have is a horde of observations in search of contextualization and deeper explanation.

Representation: How neuroimages are perceived and interpreted outside the lab

Nowadays, the greatest conceptual abstraction is to be found in conceptual images… the greatest imagination is to be found in scientific texts. Thus, behind one’s back, the hierarchy of codes is overturned. Texts, originally a metacode of images, can themselves have images as a metacode (Flusser 2000).

I turn now away from the producers of NIs and investigate how the claimed (by neuroimagers) and marketed (by press officers and mass media) visually supported arguments about the mind-brain connection compare with what is actually perceived outside the lab. Several studies have highlighted the significant weight visual evidence carries over non-visual evidence, such that an image has a greater heuristically persuasive
power and is deemed to be an even more accurate representation of a given phenomenon than are statistical and numerical presentations (Dumit 2004).

The fact that an image embodies some information does not suffice to account for its representational content. Since both referents and contents can be assigned by stipulation, just about any object can be used to stand for anything (Goodman 1976; Roskies 2008). Reiterating the technological and methodological lacunae presented earlier, Perini (Perini 2012) has pondered how NIs support scientific claims if their generation is not a simple matter of nature re-presenting itself legibly. Perini argues that NIs are not mimetic in the way photographs are since what they allegedly represent (location and level of neural activity) are not visual properties. This means that comprehending an image as a representation of something else always involves a kind of interpretation, and this representation hinges on shared interpretive practices (Perini 2012). NIs are presented not only as (at best) a substitution (i.e. a hypothetical construct) or as an epistemological or heuristic scientific observation, but rather as an actual phenomenological realization of the brain (Huber 2009).

Clearly, not only do non-professionals and neuroimagers not share the same epistemological field (which would facilitate effective communication between them), they are also separated in their epistemological status and roles (non-professional rarely can contest scientists’ claims without the assistance of other scientists). Roskies (Roskies 2010) concluded that functional NIs’ epistemic status rests on inferential distance: the actual biological phenomenon studied is inferentially far removed from the images themselves. Thus, a NI has a strong impact on the viewer while having limited scientific content. Roskies sets apart actual inferential distance (inferences inherent to the scientific procedure) from apparent inferential distance (the confidence non-professionals have in the scientific conclusion based on the NI). When these inferential distances come apart, people are prone to assign an unwarranted epistemic status to scientific claims. This means that NIs take on evidential roles not as a direct representation of natural phenomena, but only as the result of activities aimed at assigning referents and attributing content to them. These activities, in turn, are not scientifically objective and are heavily influenced by cultural and social norms, and we have to decipher and unfurl these norms if we are to see beyond the mediation of the visual (Lynch 1991; Joyce 2008; Burri 2012).

An example of a tacit scientific culture that fuels inferential distance is that NIs have by now come to constitute an aesthetic (Burri 2013; Aguirre 2014), and an affordable one at that since achieving a ”standardized” aesthetic is made easy with the availability of free statistical parametric mapping software that perform the above processes and
allows for push-button analysis of neuroimaging data with minimal understanding of the many statistical processes and assumptions (Aguirre 2014; Joyce and Hayasaka 2012; Lynch 1991). To highlight this point, Burri (Burri 2012) quotes a sales manager of an MRI scanner manufacturer who described neuroimaging conferences as follows: “It is like a beauty contest... You must see beautiful images that are high in resolution, that are luminous and perfect”.

In that regard, Frow (Frow 2012) investigated digital image processing guidelines of leading contemporary interdisciplinary science journals. These guidelines were put in place in order to detect inappropriate image manipulation after an image had been captured. Most guidelines focus on two concerns which pertain directly to the critique of functional NIs. The first requires that any adjustment made using digital processing must be applied to the whole image rather than selectively to specific parts of it. Secondly, adjustments that obscure or remove information from the original image are forbidden. Frow asserts that these guidelines reflect both a desire to redefine acceptable and unacceptable practices in image production, as well as a pressure to produce ever more visually appealing images to embellish journal covers. This approach might be interpreted as hypocritical given the statement of *Nature*’s editor in these guidelines that “beautification is a form of misrepresentation. Slightly dirty images reflect the real world”

Another interesting point raised by Frow is that these guidelines also serve the goal of protecting the scientist as a skilled professional: before the advent of digital image processing, the creation of scientific images required a substantial level of technical mastery, but the need for such expertise has long evaporated thanks to photo editing software. Thus, Frow’s study highlights a double standard: while image processing of certain data is now unacceptable in some esteemed circles due to fear of excessive manipulation, a significantly more excessive (and often not accounted for) manipulation is celebrated in other circles.

The aesthetic angle reveals a tension between what is known scientifically and what is presented publicly, but how much of this discussion is relevant to the general public? Which scientific news make it to popular media? Suleski & Ibaraki (Suleski and

---

13. A swift, back of a napkin inspection of the author guidelines of the six highest impact factor journals in the category of radiology, nuclear medicine, and medical imaging according to Thomson’s Journal citation reports (Human Brain Mapping, Radiology, NeuroImage, JACC: Cardiovascular Imaging, Circulation: Cardiovascular Imaging, and Journal of Nuclear Medicine) has shown that these journals do not address image processing in their guidelines as of the date of submission of this MS.
Ibaraki 2010) show that roughly only a permille of published scientific papers reach the mass media, with health/medicine papers taking the lion’s share of coverage. This miniscule slice from the humongous scientific output is a direct result of a phenomenon dubbed scientific sensationalism: a complex process in which very specific bits of scientific knowledge are aggressively pushed to the forefront of mass media to receive a disproportionate piece of the public’s attention. Sensationalism has several ingredients that are important for understanding the success of functional NIs. First, a conflict of interests and a bilateral miscommunication between academia and mass media as well as differences in reporting style typical of each (Ransohoff and Ransohoff 2001; Woloshin and Schwartz 2002; Rose and Abi-Rached 2013)\(^{14}\). Second is a bias shared by both acade me and mass media for publishing predominantly positive results while omitting negative ones, thus skewing both scientific and public perception (Easterbrook et al. 1991; Koren and Klein 1991; Cassels et al. 2003; Zuckerman 2003; Caulfield 2005; Brechman, Lee, and Cappella 2009; Gonon, Bezard, and Boraud 2011)\(^{15}\). This bias is well documented in functional NI as well (Ioannidis 2005b; John, Loewenstein, and Prelec 2012; Vul and Pashler 2012; Ioannidis et al. 2014) Third, mass media tend to flatten scientific reports and strip them of the many complexities that characterize them (Woloshin and Schwartz 2002; Beck 2010; Schwartz et al. 2012). Combined, these phenomena generate hype fluctuations that misinform the public and cause it to doubt scientific results (Ransohoff and Ransohoff 2001), and produce errant cultural residues (Conrad 1997; Gonon, Bezard, and Boraud 2011).

Evidently, neuroscience and neuroimaging, being part of the health world are no exception and experience sensationalism (O’Connell et al. 2011) and aggressive commercialization too (Chancellor and Chatterjee 2011). Barring highly irregular cases, the flow of information from lab to media is characterized by the ironing out of technological and methodological concerns, the discard of cautionary comments and the omission of alternative explanations, all resulting in distorted conclusions committed to popular memory. If these were merely misconceptions rampant within the community

\(^{14}\) As an example, Robillard & Illes (Robillard and Illes 2011) report that nearly half of the neuroscientists they have interviewed claimed that their academic institutions frown upon their efforts to communicate their research to the public.

\(^{15}\) Sensationalism can escalate the commercialization of academia (Downie and Herder 2007; Hong and Walsh 2009) and even bias scientific practice: fMRI studies were cited three times more often than lesion studies of the same brain region, mostly due to the fact that they were published in higher impact factor journals (Fellows et al. 2005).
of neuroimagers it would amount to scant scientific practice. The problem is different: while they may know better, what they communicate to the outside is not. That equals unethical practice.

An example of scientific sensationalism at the academic level can be seen in Charest et al. (Charest et al. 2014) fMRI study of semantic space. In the Significance section of the article, the authors write “our results demonstrate that fMRI has the power to reveal individually unique representations of particular objects in the human brain. The novel method might help us understand the biological substrate of individual experience”. However, in the concluding paragraphs of their discussion they write “It is important to note that the predictions of perceptual idiosyncrasies from the hIT representation, although robustly better than chance, are not very precise. Precision estimates depend on many factors, and have little meaning beyond the context of a particular study”.

An example of scientific sensationalism in mass media is a famous article published in The New York Times on 11/11/2007 titled “This is your brain on politics”. In anticipation of the 2008 US presidential elections, the article described an fMRI experiment in which twenty swing voters were scanned while images and videos of candidates were presented to them. The authors concluded that the brain responds differently to Republican and Democrat candidates as well as to the words “republican” and “democrat” themselves. They even ventured that voters had mixed feelings toward Hilary Clinton, while Mitt Romney showed potential. The article generated instantaneous political and cultural fervor, and that of the scientific community soon followed as it had realized that this attempt to impact the results of the elections was not only egregious in intent, but also flagrantly scientifically vacuous. One of the leading cognitive neuroscientists, Russell Poldrack commented:

It was really closer to astrology than it was to real science… it epitomized everything that a lot of us feel is wrong about where certain parts of the field are going, which is: throw someone in a scanner and tell a story about it… people will start to see fMRI as neophrenology, just telling stories and not giving explanations (Ramani 2009).

The elections story becomes even more nefarious when considering a meeting in 2005 that brought together leading neuroscientists, ethicists, and journalists to discuss various aspects of neuroimaging. Participants were of the opinion that neuroscientists have a

---

responsibility to clarify to the public what are the limitations of cognitive neuroscience research. Some even argued that scientists are obliged to be more rigorous in their research (Check 2005).

The elections study has led researchers to investigate directly the effect NIs have on layperson. This work has yielded conflicting results: some argued that part of the credibility afforded to brain imaging lies in the image itself (McCabe and Castel 2008; Roskies 2008; Keenhe and Fischer 2011; Ikeda et al. 2013; Saks et al. 2014), while later work reported no such special effect (Gruber and Dickerson 2012; Hook and Farah 2013; Michael et al. 2013; Schweitzer, Baker, and Risko 2013). Interviews conducted with neuroimagers, patients, and other non-professionals show that NIs are performative and enact the body rather than transparently represent it (Joyce 2005; Casini 2011). The performative can also easily cross the line and become deceptive, as evidenced by quotes from MRI technicians who noted that “It’s easy to tweak the parameters to make something that’s not there” or admitted that MRI images are all “smoke and mirrors” (Joyce 2005). Burri (Burri 2013) unravels similar observations from her interviews: when initially asked to describe MRI images, scientists portray them as a “document which depicts reality 1:1”; “it’s a photograph”; or “if you would slice the body, it would look exactly like that”. The flip side is given via a psychiatric patient who had undergone fMRI and stated that “It’s a picture of who you really are. On the inside” (Cohn 2010). Burri (Burri 2013) quotes a patient stating that “the image is... something irrevocable”, echoed by a physician who said that “The images persist. Patients remember them well”. Another neurophysiologist recounted presenting at a conference and admitted “We presented [the images] in a really wrong way. [The audience], however, liked it. People were not aware that the images were wrong”. However, when asked for a minimal level of reflection, the quotes from Burri’s interviews quickly change to “there is a danger in the images. Because images sometimes suggest more than they should”, or “in every image there is something delusive”. Alarmingly, that is not the full extent of it. Burri quotes a radiology professor claiming that “images pretend a lot of authority, seeming authority that absolutely doesn’t exist”. This state of affairs has led Carp (Carp 2012a) to claim that “A motivated researcher determined to find significant activation in practically any brain region will very likely succeed – as will another researcher determined to find null results in the same region”.

Expressing his opinion of the state of neuroscience, Martell (Martell 2009) was very adamant in stating that “the ability of neuroscientists to use neuroimaging reliably to predict (and perhaps… postdict) thoughts or behavior is currently nil”. Wolpe (Wolpe 2006) stated that “science has become one of the most powerful and pervasive forces
for change in modern societies. As the professionals at its helm, scientists have a unique responsibility to shepherd that change with careful ethical scrutiny of their own behaviour and thoughtful advocacy of scientific research”. Similarly, Lavazza & De Caro (Lavazza and De Caro 2010) assert that “When one comes to the issue of human agency, great caution should be used before drawing bold philosophical, political, and social conclusions from neurological findings, whose correct interpretation and value are still extremely controversial”. To add insult to injury, research shows two worrisome trends in public’s perception of the scientific explanation of mental illness. First, the discourse is dominated by biological models. Second, and as an upshot, reductionist causal explanations tend to exacerbate negative feelings toward the mentally ill (Angermeyer et al. 2011; Schomerus et al. 2012; Kvaale, Gottdiener, and Haslam 2013; Lebowitz and Ahn 2014). This exact public reaction has been documented previously in the case of genetic explanations in general and those of brain/mind in particular (Dar-Nimrod and Heine 2011; Haslam 2011). Rectifying the epistemological limitations of fMRI is an ethical imperative on two levels: firstly, because they constitute a threat to the quality of research (Kaposy 2008; Anderson, Mizgalewicz, and Illes 2012; Bluhm 2013; Peterson 2003); and secondly (and perhaps more importantly) because they perpetuate a skewed misconception of what brains are and do as well as what their connection is to cognition, mind, and the self.

Nonetheless, it must be remembered that the public, though not well versed in the particulars of neuroimaging, is far from inert: people exercise critical judgment of scientific news and incorporate various discourses to form an opinion (Wynne 2001; Meurk et al. 2014). Furthermore, the public must not be seen as a monolithic entity, but rather as enjoying different perceptions of science, brain, and mind (Cohn, Dumit, and Roepstorff 2003). However, a lingering question is have audiences already been numbed by NIs in popular media? Whitely (Whiteley 2012) examined mass media publications between 2005 and 2009 and learned that coverage of fMRI studies has often substituted NIs with artistic renditions. The question is whether NIs have thus become synonymous to brain and cognition or have they become void of meaning beyond their aesthetic value and reduced to a typological brand name. This visual saturation scenario might explain why studies published from 2013 onward failed to support the argument the NIs impact viewers in a unique fashion. Another explanation comes from Fernandez-Duque et al. (Fernandez-Duque et al. 2014), who show that neuroscientific explanations (be they accompanied by NIs or not) carry a distinct allure bestowed with exaggerated credence which amounts to a conceptual bias. Vidal (Vidal 2009) provides a detailed anthropological account of this phenomenon and dubs it the brainhood, or the cerebral subject ideology.
In conclusion, it is safe to say that within the scientific community, neuroimagers reject NIs as a visual technology and categorize it instead as a numerical representation. This faux-naïf attitude allows neuroimagers to manipulate two worlds simultaneously: by adhering to mathematization they gain scientific credence; while by exorbitantly processing colorful images they mesmerize non-professionals. As a corollary, neuroimagers gain support on all fronts (Beaulieu 2002). The persuasive power of NIs is abused as a communication tool to promote specific scientific ideology at the expense of rivaling theories (De Vos 2014). This combination of treacherous images, a cognitive bias toward visual imagery, and a conflicted scientific culture hidden from the public eye can quickly turn NIs to an ethically loaded gun.

An Alternative Account of Cognition: Emergence and Evolution

It seems incumbent to conclude the criticism offered here by proposing an alternative account of the story behind the brain-mind connection. To that end, I return to the critique of localization and ponder that while there is no doubt that reductionism has served science extremely well, does this success necessarily translate into a monopoly over fundamental ontology? Is reduction the best tool now that we know the brain is a complex, multi-player, multi-layered, spatiotemporally spread, non-linear, and heavily context-dependent system (e.g. (Ellis 2009))? What could be a more philosophically as well as physiologically apt interpretation of the link between the brain and the mind? To accept both that there are physiological underpinnings of cognition, emotions, and behavior, as well as that those phenomena exist only at the organismal level and above, forces us to identify and characterize a process by which electrochemical signal transduction is transformed into abstract, intangible, and vaporous thoughts and feelings. Since we currently lack the knowledge to offer any particular property as such a Holy Grail psychophysical compiler, I offer here (not as novelty) as candidates neuronal emergent properties.

Emergent properties are novel traits of a system that result from unique spatiotemporal relational dynamics of the building blocks of lower organization levels of that system, which are irreducible to the principles governing those lower levels (Rueger 2000; Silberstein 2002; Newman 2011). In this fluid interplay, building blocks may constrain - but not determine - the attributes of higher-level traits, such that the phenomenon generated at each level obeys the rules of that level, not of lower, and that

17. Emergence is often seen as an opposition to reductionism on account of irreducibility (Delehanty 2005).
higher levels may exert regulatory power over lower levels but not vice versa (Newsome 2009).

In brains, the wide range of neuronal elements, along with their contextual spatiotemporal orchestration of connectivity patterns and interactions in a complex 3D architecture, combine to generate novel processes that should be considered an emergent property of the nervous system as it complexifies and goes up organization levels. To date, a host of emergent properties have been documented, covering many aspects of neuronal operation, such as cortical oscillations (Wang 2010; Whittington et al. 2011), cortical coordination dynamics (Bressler and Kelso 2001), spiking rate (De Sancristóbal et al. 2013), circadian rhythms (Muraro, Pírez, and Ceriani 2013), electrical properties of dendritic spines (Yuste 2013), pre/postsynaptic terminal structure (Emes and Grant 2012), formation of neuronal assemblies (Fingelkurts, Fingelkurts, and Kähkönen 2005), synchronization of neuronal assemblies (Lindsey et al. 1997), functional information segregation (Ma et al. 2014), synaptic current and phase locking (Deco et al. 2008), and cognition in general (McIntosh 2000).

The idea here is that, given that brain activity is no stranger to emergent properties, higher-level neuronal actions combine to create the necessary conditions for the emergence of a quasi neuronal-independent process that utilize the dynamic interactions with the body and the environment to create specific portions of what we call the mind.

If we were to accept emergence as an unspecified mechanism shaping brains and cognition, then a next step is to accept that the system in question had undergone an evolutionary change (a change can be ontogenic only, but then it is of no consequence to the species). By agreeing that an evolutionary account is appropriate here, we then need to turn our attention to possible evolutionary mechanisms at play. Bunge (Bunge 1977) argued that radical novelties emerge out of previously existing things, such that emergence and levels of organization are dynamic orders and are features of an evolutionary process. In biological terms, what Bunge referred to, later became known as exaptation (Gould and Vrba 1982): the process by which features that now enhance fitness but were not built by natural selection for their current role have appeared.

Exaptation has already found its way into neuroscientific thought via neuronal re-use hypotheses, which argue that the complexification and evolution of the brain takes place via preservation, extension, and combination of existing networks (e.g. (Sporns and Kötter 2004; Just and Varma 2007)). The most sophisticated neural re-use theory is the massive redeployment hypothesis (Anderson 2007). It argues that existing components, which serve a specific purpose, are recruited for new purposes and combined to support new capacities without disrupting their previous functionality. Thus, each element does
Shifferman

only one thing, and it continues to do it regardless of the networks and complexes it is part of. It is far beyond the scope of this paper to develop the evolutionary exaptationist model of the transition from brain to mind, and it is presented here only as an abstract appetizer of a prolegomenon, a probable non-reductive account of the link between nervous systems and cognition, standing in opposition to the materialist localization project defended by neuroimagers.18

Conclusion

I argued here that the use of functional neuroimages for the purpose of supporting localization claims embedding the mind and consciousness exclusively in well-defined brain regions is an unethical scientific endeavor. By adopting a four-prong critique, I challenged the validity of the basic tenets and practices of neuroimaging at the technological, methodological, and philosophical levels. This criticism highlights lacunae that result in five degrees of separation between the biological phenomenon of neural response to a cognitive task and the NIs that allegedly represent it. I then examined the psychological impact of NIs and learned that they are borne out of a scientific culture with a strong penchant for aesthetics, a bias kept hidden from the public eye. The resultant hypothesis is that cognitive neuroscience’s use of functional NIs is unethical by knowingly allowing flawed conclusions to trickle down from labs to policy makers, mass media, and the public, thus skewing public understanding of the fundamental issue of the mind-brain connection and cementing an erroneous interpretation of this problem. I am of the opinion that a scientific culture that cultivates and celebrates aesthetics over scientific accuracy and reliability; that knowingly disseminates distorted data masked by the appeal of heavily engineered images should not be surprised by fiascos such as the 2008 US presidential elections fMRI scan study. To pretend to be outraged by it is to turn a hypocritical blind eye to the academic climate that facilitated it while working toward perpetuating the very same problems that constitute a hurdle in the quest for a candid scientific effort to understand the mind and its possible neuronal basis.

18. Evolutionary theory will also allow further development of the distinction between neurotype and cognitype, by distinguishing between two possible candidates for selection processes.
References


Ball, Tonio, Johanna Derix, Johanna Wentlandt, Birgit Wieckhorst, Oliver Speck, Andreas Schulze-Bonhage, and Isabella Mutschler. 2009. “Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative...


Shifferman


Shifferman, Eran M. 2011. *The evolution of quantity estimation in the animal kingdom*, Faculty of Humanities, School of Philosophy, Tel Aviv University, Tel Aviv.


