Severe Tests in Neuroimaging: What We Can Learn and How We Can Learn It

M. Emrah Aktunc*†

Considerable methodological difficulties abound in neuroimaging, and several philosophers of science have recently called into question the potential of neuroimaging studies to contribute to our knowledge of human cognition. These skeptical accounts suggest that functional hypotheses are underdetermined by neuroimaging data. I apply Mayo's error-statistical account to clarify the evidential import of neuroimaging data and the kinds of inferences it can reliably support. Thus, we can answer the question "What can we reliably learn from neuroimaging?" and make sense of how this knowledge can contribute to novel construals of cognition.

1. Introduction. Considerable methodological difficulties abound in neuroimaging; several philosophers of science, notably Klein (2010) and Roskies (2008, 2010), have called into question the potential of neuroimaging studies to contribute to our knowledge of human cognition. One general conclusion in these skeptical accounts is that functional hypotheses about cognitive processes are underdetermined by neuroimaging data. Yet, functional neuroimaging research continues to grow, so there is a need to address the question of what it is that we can learn from neuroimaging. After briefly discussing works by Klein and Roskies, I will apply to functional magnetic resonance imaging (fMRI) Mayo's error-statistical (ES) notions of severe tests, error probabilities, and a hierarchical framework of models of inquiry.[1] The ES account helps clarify the evidential import of neuroimaging data and formulate the conditions under which we can reliably infer that we have evi-

*To contact the author, please write to: Ozyegin University, Nisantepe Mah. Orman Sok. Cekmekoy, Istanbul 34794, Turkey; e-mail: emrah.aktunc@ozyegin.edu.tr.

†I wish to extend my gratefulness to Deborah Mayo, Richard Burian, Aris Spanos, and Lydia Patton for their great help and advice in the formation of the ideas presented in this paper.

1. Parallel ES analyses of other neuroimaging media can be carried out mutatis mutandis.

Philosophy of Science, 81 (December 2014) pp. 961–973. 0031-8248/2014/8105-0021$10.00
Copyright 2014 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/677691 Published online by Cambridge University Press

dence for or against functional hypotheses. Thus, we can answer the question "What can we reliably learn from neuroimaging?" and make sense of how this knowledge can contribute to cognitive neuroscience and lead to novel construals of cognition.

2. Skepticism about Functional Neuroimaging. Klein (2010) and Roskies (2008, 2010) have different arguments based on various premises, but they both come to similar skeptical conclusions regarding the epistemic value of neuroimaging. Klein (2010) has raised criticisms directed at the use of statistical hypothesis testing in neuroimaging experiments. When researchers compare observed brain activation in a control condition against an experimental condition, where subjects perform the given cognitive task, they test the null hypothesis that there is no significant difference between the conditions against the alternative hypothesis that predicts a difference. The null hypothesis assigns probabilities to certain outcomes in the scenario where it is true, and the probability of a certain outcome under the null hypothesis is its *p*-value. If the *p*-value of the observed outcome is smaller than a predetermined significance threshold, then we have a significant result; we reject the null hypothesis and conclude that there is a significant difference between the control and experimental conditions. The central premise in Klein's argument is that in neuroimaging it is relatively easy to find significant results even when there is no real effect. For example, in order for a region of the brain to be identified as 'active', there has to be a statistically significant difference between degrees of observed activation in that brain region across control and experimental conditions, so choosing an overly liberal threshold for significance may yield spurious results. The charge is that when we observe significantly high activation in a given brain region, this may not be because there really is increased task-related activity in that region, but because we have chosen a significance threshold so liberal that it picks up background noise as if a real effect. Indeed, this is a real problem and is known in the ES literature as the simple fallacy of rejection.

Of course, there are factors other than the chosen threshold that may bias analyses and yield significant results in the absence of a real effect. Klein (2010) discusses how the signal-to-noise ratio in neuroimaging can be improved by increasing the number of subjects, which increases the sensitivity of the experiment. Consequently, significant results yielded by the experiment may have occurred only because the number of subjects was increased and not because there is a real effect. Klein claims that neuroimaging runs into problems not as a consequence of its inherent characteristics but because it requires statistical hypothesis testing to draw inferences about functional hypotheses. Because of these and other similar problems, Klein concludes that the best use of neuroimaging experiments is to serve as "first-pass sanity check[s] on experimental data," which can never "confirm functional

https://doi.org/10.1086/677691 Published online by Cambridge University Press

hypotheses" (2010, 275). Klein is obviously right in criticizing flawed practices of hypothesis testing. However, it is crucial to note that these flawed practices occur not because of any inherent characteristics of statistical hypothesis testing but rather because of misunderstandings and misuses of these tests. The ES notions of severe tests and error probabilities help remedy these misunderstandings and misuses, as is described below.

Roskies has also emphasized methodological issues in neuroimaging and suggested that the perceived epistemic status of neuroimages is higher than their real epistemic status (2008, 2010). In order to interpret neuroimaging results correctly, we need to have a way of knowing their actual epistemic status. According to Roskies, "Determining actual epistemic status will involve a characterization of the inferential steps that mediate between observations and the phenomena they purport to provide information about" (2010, 197). She introduces the term *inferential distance* to refer to the totality of these inferential steps and argues that the problem in neuroimaging is the mismatch between the "actual inferential distance" and the "apparent inferential distance." There are a great number of technical and inferential procedures in neuroimaging that have to be carried out between initial measurements of brain activation and final neuroimages. Because of the complexity of these procedures, Roskies concludes that the nature of the inferential steps in neuroimaging cannot be sufficiently characterized, which lowers the reliability of inferences drawn from neuroimaging data.

Roskies is definitely right that "neuroimages are inferentially distant from brain activity" (2008, 30) and results are too often overinterpreted. But her further conclusion that the inferential distance in neuroimaging cannot be univocally characterized is questionable. Roskies's inferential distance problem can be satisfactorily addressed when we apply a hierarchical framework of models of inquiry to neuroimaging. In this framework, we can break down a neuroimaging study into its component parts from experimental design to initial data collection, and from preprocessing of raw data to statistical modeling and analysis. We can then assess the error characteristics associated with each component. Let us say that an experiment yields significantly higher activation in a certain brain region in the experimental condition. Thus, researchers conclude that the cognitive task performed by subjects led to this result, and they draw the substantive inference that the brain region with higher activation is involved in the performance of this cognitive task. Now, we can carry out ES analyses of this experiment looking carefully into the error characteristics or error probabilities of its component parts. If these error probabilities are high, then we do not have strong evidence for the researchers' conclusion. This is because the significant result may have been obtained owing to a bias introduced by a component part of the experiment. For example, the neuroimaging scanner may have been oversensitive and detected background noise as task-related activation, or, as Klein pointed out, the

https://doi.org/10.1086/677691 Published online by Cambridge University Press

chosen significance threshold may have been too liberal. If the error probabilities associated with the components of an experiment are low enough to rule out or minimize errors, then we can safely conclude that researchers have support for their substantive inference. This is how we can go the inferential distance, as it were.

3. An Error-Statistical Approach to Functional Neuroimaging. Klein and Roskies rightly emphasize methodological issues in neuroimaging, but they both seem to assume that these cannot be satisfactorily resolved. This can be contested because Mayo's ES notions of severe tests and error probabilities can be employed to address the issues raised by Klein and Roskies and problems of inference in neuroimaging can be resolved. In order to tackle problems of inference, we need to know the conditions under which we can conclude that we have evidence for a hypothesis. We can start with a weak requirement: for an experiment or test, the weakest requirement is that it should not be guaranteed to find evidence for some effect regardless of whether or not there is a real effect. This gives us what Mayo and Spanos (2011) call the *weak severity principle*: data **x** do not provide good evidence for a hypothesis H if **x** results from a test procedure with a very low probability or capacity of having uncovered the falsity of H (even if H is incorrect). This notion is the fundamental basis of the account that scrutinizes experiments by analyzing them with respect to their error probabilities—this is what Mayo calls the ES account. Error probabilities provide the information on how frequently methods of research can discriminate between alternative hypotheses and how reliably they can detect errors.

In light of these concepts, we can better address the question "When do data **x** provide good evidence for a hypothesis H?" To do this, we can take Mayo's *full severity principle* as a guide, which states, "Data **x** (produced by process G) provide a good indication or evidence for hypothesis H (just) to the extent that test T severely passes H with **x**" (Mayo 2005, 100). For a hypothesis H to pass a severe test T with **x**, two things must obtain: first, data **x** fits or agrees with H, and second, test T would have produced, with high probability, data that fit less well with H than **x** does, were H false (Mayo 1996, 2005). Thus, data **x** are evidence for hypothesis H just to the extent that the accordance between **x** and H would be difficult to achieve were H false. One must have done a good job at probing the ways one may be wrong in inferring from an accordance between data **x** and hypothesis H to an inference to H (as well tested or corroborated). It is important to note that the severity of a test is not a feature of only the test itself. Severity assessments are carried out always on a specific test T, with specific test result $\mathbf{x}_0$ and a specific hypothesis H, so severity is a function of three things: the test (or the experiment), the data, and the specific hypothesis about which an inference is drawn (Mayo 2005). We can use the abbrevi-

https://doi.org/10.1086/677691 Published online by Cambridge University Press

ation SEV($T,\mathbf{x}_0,H$) to mean "the severity with which H passes test T with $\mathbf{x}_0$" (Mayo and Spanos 2011), where the severity function SEV($T,\mathbf{x}_0,H$) can be calculated to get a quantitative value between 0 and 1. The notion of severity can also be employed in discussing error characteristics of experimental tools. In any given experiment, to assess whether or not it constitutes a severe test of the hypothesis of interest, we need to know the error characteristics associated with its components, such as instruments and techniques used for collection and processing of data, and statistical modeling and analyses.

In fMRI, often the hypothesis of interest, which I call the real effect hypothesis, predicts higher activation in some brain region in response to a cognitive stimulus or task. We can assess any given experiment with respect to whether or not it has put a specific real effect hypothesis to a severe test. If it has not, then we evaluate its results accordingly. If an experiment of low severity yields a result fitting the real effect hypothesis of interest, this does not constitute evidence for the hypothesis, because, with high probability, the experiment would have yielded a fitting result even if the hypothesis were false. The ES notions of severity and error probabilities help us properly assess the epistemic value of fMRI studies that may suffer from problems discussed by Klein, such as liberal significance thresholds or overly large samples. With high probability, an experiment with an overly large sample may yield results fitting the real effect hypothesis even when it is false. Such an experiment constitutes a low severity test of the hypothesis, and we would not make the mistake of taking its results as evidence for it. The problem arising from increasing the number of subjects is called the large-N problem in philosophy of statistics, which can be resolved by complementing statistical tests with Mayo's notion of severe tests. As N gets larger, the variance of the data is reduced. Since the variance of the data is the denominator in the calculation of the test statistic, mathematically, the observed test statistic gets larger independently of the truth or falsity of the alternative hypothesis. Consequently, as N gets larger, it becomes more probable to obtain a significant result in the absence of a real effect, that is, when the null hypothesis is true. Therefore, the ES account differentiates between experiments with different numbers of subjects; a significant result is less indicative of a real effect if it was obtained in an experiment with a large sample than in an experiment with a smaller sample (Mayo 1996, 2005). To illustrate quantitatively, here is how values of severity would change with the same outcome, $d(\mathbf{x}_0) = 1.96$, $\sigma = 2$, and for the same inference, $\mu > 0.2$, but on the basis of different sample sizes, n: for $n = 50$, SEV($\mu > 0.2$) = .895; for $n = 100$, SEV($\mu > 0.2$) = .831; and for $n = 1000$, SEV($\mu > 0.2$) = .115 (Mayo and Spanos 2011). As can be seen, the severity of inferring $\mu > 0.2$ with outcome $d(\mathbf{x}_0) = 1.96$ decreases as the sample size n gets larger. In fMRI, we can think of the outcome as significantly increased activation in a brain re-

https://doi.org/10.1086/677691 Published online by Cambridge University Press

gion as predicted by the real effect hypothesis. If such an outcome obtains in an experiment with an overly large sample, we do not take it as evidence, because the experiment would be a low severity test of the hypothesis. We may have obtained that outcome as a result of having a large sample size that makes for an oversensitive test picking up noise as if a real effect. In this way, we can resolve problems of inference resulting from overly large samples in fMRI studies. Other problems Klein raised can be similarly addressed when we employ ES notions. For example, we can show that the more liberal the significance threshold in an experiment, the lower is the severity of that experiment as a test of the hypothesis of interest, because the overly liberal threshold would increase the probability of obtaining a significant result in the absence of a real effect.

The ES account can also be employed to address Roskies's inferential distance problem. As stated above, the severity of a test is a function of three things: the experiment, the data obtained in the experiment, and the specific hypothesis. In order to assess severity, one needs accurate characterizations of all these aspects of an experiment. To achieve this, we can look at an experimental inquiry in terms of a hierarchy of models that connect the primary hypothesis or question being investigated to the detailed procedures of data generation and analysis. For any experimental inquiry, three types of models are defined: models of primary hypotheses, models of experiment, and models of data (Mayo 1996). These models help clearly describe the local procedures that are required to establish the connection between raw data and the hypotheses of interest.

The primary model includes the local hypotheses, which may have been derived from a higher-order scientific theory or from previous studies, and they correspond to a given primary question or problem. In fMRI, the primary hypothesis often predicts the amount of activity in a given brain region. The experimental model provides the link between the data and the primary hypothesis being tested. Mayo talks about two functions of the experimental model. The first function is to provide "a kind of experimental analog of the salient features of the primary model" (Mayo 1996, 133). If the primary problem is testing a hypothesis, then the experimental model tells us what is expected to obtain in this experiment if the hypothesis is true, possibly by using other auxiliary hypotheses. The second function of the experimental model is "to specify analytical techniques for linking experimental data to the questions of the experimental model" (Mayo 1996, 134). Because of many sources of error that influence the data collection process, the data will very rarely—perhaps never—agree exactly with the experimental prediction. In this case, the experimental model may statistically formulate the link between the primary hypothesis and the data model. The data models provide the answers to two types of questions: the before-trial question is about how raw data should be collected and modeled to be put in

https://doi.org/10.1086/677691 Published online by Cambridge University Press

canonical form in order to be linked to the experimental model; the after-trial question is about how we can check whether or not the data collection procedures were in line with the assumptions of experimental models.

The application of the hierarchical framework of models to fMRI can help delineate and appraise the experimental procedures that are involved. We can also find out to what extent, if at all, these procedures may introduce errors and influence the results independently of the truth or falsity of functional hypotheses, so that we can assess the reliability of the inferences drawn on the basis of those results. Thus, this framework provides a way of characterizing the steps in the inferential distance between raw data and final neuroimages. We can start by looking at the kinds of hypotheses that fMRI experiments are meant to test. Often, these are statements derived from theories of cognition or results of previous studies and take the form "brain region X is involved in the performance of cognitive task C." From this functional hypothesis we can derive the prediction that "when subjects perform C, there is going to be a significant amount of activation in brain region X." Along with the functional hypothesis, this prediction can be placed in the primary model. In fMRI, the connection between the primary hypothesis and its experimental analogs can be made by stating what would happen in the experiment if the primary hypothesis is true: "as subjects in the fMRI scanner perform an example of the cognitive task C, the scanner will register a significantly high amount of activation in brain region X."

Designing the experiment, choosing the experimental task, and other related aspects would be placed in data models. The experimental design ensures that we get the kind of data to test the primary hypothesis; hence, it helps connect the actual experiment to experimental models and primary hypotheses of the inquiry. Once we have raw data, we have to put them in canonical statistical form in order to carry out statistical analyses, which is not easy because fMRI scanners yield extremely complex data sets. Decisions about preprocessing steps such as the use of spatial and/or temporal filters, signal averaging, and any other necessary procedures would be placed in data models. Once preprocessing is complete, we have to obtain a statistically adequate model of the data generating mechanism and then carry out significance tests to make inferences about the primary hypothesis. All these procedures would also be placed in the data models. The application of this hierarchical framework provides a more complete understanding of how an fMRI experiment works and what it can and cannot give us. Once we identify and correctly place the component parts of an experiment in this framework, we can then assess the error probabilities, or error characteristics, associated with each component. Thus, we can find out how and to what extent they may introduce errors in the experiment. Naturally, fleshing out the details of how fMRI works and how these error probabilities can be assessed is beyond the scope of this paper. I have provided in-depth

https://doi.org/10.1086/677691 Published online by Cambridge University Press

treatments of these topics elsewhere (Aktunc 2011, 2014). Here I focus on a further elaboration of what goes in the primary model of an fMRI experiment.

4. Kinds of Hypotheses in fMRI. As seen above, a statement of the form "Brain region X is involved in the performance of cognitive task C" is placed in the primary models as the hypothesis to be tested in an fMRI experiment. However, this statement is ambiguous; one has to consult some basic characteristics of fMRI to disambiguate it. The goal in fMRI is to relate changes in brain physiology over time to an experimental manipulation. This relation is established in terms of the hemodynamic response, also known as the blood-oxygenation-level-dependent (BOLD) response. If a certain brain region is involved in the performance of a cognitive task, then, when an individual performs it, there is going to be increased activation in that region of the brain. Increased activation causes an increased need for energy, and this leads to an increase in local glucose metabolism and oxygen consumption. Oxygen is carried to cells by oxygenated hemoglobin in the blood, so there results an imbalance between concentrations of oxygenated and deoxygenated hemoglobin in activated regions. Since these two types of hemoglobin have different magnetic properties, the imbalance leads to inhomogeneities in the magnetic field. The fMRI scanner detects these inhomogeneities and thus provides data on hemodynamic activity in terms of the magnitude of the BOLD response (Huettel, Song, and McCarthy 2008).

Huettel et al. (2008) state that in a typical fMRI study three distinct kinds of hypotheses are involved:

- *Hemodynamic hypothesis*: a hypothesis about relationships between hemodynamic activity and performance of cognitive tasks.
- *Neuronal hypothesis*: a hypothesis about neuronal activity. Since fMRI does not directly measure neuronal activity, we cannot know for certain the amount of neuronal activity; researchers assume that if the observed hemodynamic activity is high, then neuronal activity is also high.
- *Theoretical hypothesis:* essentially a hypothesis about cognitive function. Researchers use fMRI results to address questions about how cognitive processes work and how they are realized by which neural structures and processes.

A crucial question is, which of these three kinds of hypotheses can we put to severe tests in an fMRI experiment? Since fMRI gives us data only on hemodynamic activity, it appears that hypotheses about hemodynamic activity can, at least potentially, be put to severe tests in fMRI experiments.

https://doi.org/10.1086/677691 Published online by Cambridge University Press

These hemodynamic hypotheses can be formulated in terms of statistical hypotheses, which can then be tested against fMRI data. Researchers want to see what differences there are, if any, between amounts or patterns of activation in certain regions of the brain across control and experimental conditions. Often, the hemodynamic hypothesis predicts higher activation in the experimental condition. We can use μ_0 to designate mean hemodynamic activation (in a certain brain region) in the control condition, and we can use μ_1 to designate mean hemodynamic activation (in the same region) in the experimental condition when subjects perform the chosen cognitive task. These two means, μ_0 and μ_1, as well as the difference between them, $\mu_1 - \mu_0$, are unknown parameters, and researchers are interested in making inferences about the difference $\mu_1 - \mu_0$. Statistical hypotheses about the difference between μ_0 and μ_1 can be stated as null and alternative hypotheses in a significance test. In statistical hypothesis tests, we can test the null hypothesis, H_0: $\mu_1 - \mu_0 = 0$, versus the alternative hypothesis, H_1: $\mu_1 - \mu_0 > 0$, which are formulated in the context of a statistical model of fMRI data. The alternative hypothesis so formulated is connected through a statistical model of data to the hemodynamic hypothesis, which predicts a significantly higher amount of activation in a certain brain region when subjects perform the chosen cognitive task. In an fMRI experiment, the real effect hypothesis referred to before is no other than the hemodynamic hypothesis. The critical point is that, using error probabilities, we can find out whether or not specific fMRI experiments constitute severe tests of specific hemodynamic hypotheses. Thus, fMRI data do have evidential import for hemodynamic hypotheses.

The neuronal and theoretical hypotheses are not subjected to severe tests by fMRI. The neuronal hypothesis makes an assertion about neuronal activity, and we cannot reliably infer much about that from fMRI data. For the observed hemodynamic activity could be due to factors other than neuronal activity, such as the activity of glial cells (Huettel et al. 2008). Logothetis (2008) reinforces the point that observed hemodynamic activation does not always mean that it was caused by neuronal activity. Therefore, the neuronal hypothesis could be false, and yet the fMRI experiment would not detect this. As for the theoretical hypothesis, fMRI obviously does not test for the existence of cognitive modules or functions as defined by theories of cognitive science. This is related to the criticisms of neuroimaging raised by Uttal (2001) and Hardcastle and Stewart (2002). They have argued that there is an inherent circularity in assuming the existence of localized and well-defined cognitive modules prior to doing an fMRI experiment and then taking the results of the experiment as support for modularist conclusions. When Uttal (2001) and Hardcastle and Stewart (2002) raise the circularity problem and when Klein (2010) calls into question the evidential value of fMRI data for functional hypotheses, they are talking about prob-

https://doi.org/10.1086/677691 Published online by Cambridge University Press

lems of using neuroimaging results as support for theoretical hypotheses. The ES approach shows that fMRI data cannot be taken as evidence for these types of theoretical hypotheses, but at the same time, it helps identify what we can reliably take as the fMRI finding, specifically whatever the hemodynamic hypothesis states in an experiment provided that the related inference to H_1: $\mu_1 - \mu_0 > 0$ passes severely with the data obtained in that experiment. This hemodynamic finding can then be used as an additional constraint on theories of cognitive science. Figure 1 summarizes the conclusion of the discussion above. The link between the hemodynamic hypothesis and H_1: $\mu_1 - \mu_0 > 0$ shows that the hemodynamic hypothesis is embedded into a statistical model of the data and is framed in terms of the parameters μ_0 and μ_1.

5. Hemodynamic Findings and Novel Construals of Cognition. I have argued above that the hemodynamic hypothesis need not be underdetermined by fMRI data, or at least the corroboration that hemodynamic hypotheses gain from fMRI experiments is not shaken by the underdetermination neuronal or theoretical hypotheses may suffer. In response, one may be compelled to ask, what is hemodynamic knowledge good for? In addition to providing novel constraints on existing theories, hemodynamic findings are helpful in the development of new models of cognition. I wish to illustrate this with a set of hemodynamic hypotheses well known as the hemispheric encoding/retrieval asymmetry (HERA) model.

The HERA model began as a straightforward description of empirical regularities found in positron emission tomography (PET) studies of memory; although it works differently from fMRI, PET, too, provides measurements of cerebral blood flow, so it can support hemodynamic hypotheses. Researchers had obtained differential activation patterns in left and right pre-

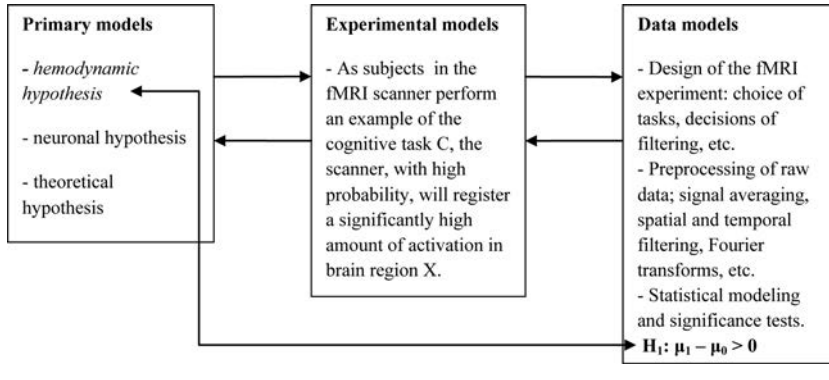


Figure 1.

https://doi.org/10.1086/677691 Published online by Cambridge University Press

frontal cortical regions when subjects engaged in encoding and retrieval tasks of episodic memories. Here I am interested in observed relationships between hemodynamic activity and performance of operationally well-defined cognitive tasks. In the context of HERA, the terms 'semantic memory' and 'episodic memory' can be used simply to refer to these different kinds of tasks without having to worry too much about the theoretical baggage these terms may carry. The HERA model was proposed in a data-driven manner; in a series of PET experiments, subjects performed three types of tasks: semantic memory retrieval, episodic memory encoding, and episodic memory retrieval. Tulving et al. (1994) summarized regularities in observed patterns of brain activation: (1) left prefrontal cortical regions are activated in semantic memory retrieval to a greater extent than right prefrontal cortical regions, (2) left prefrontal cortical regions are activated in encoding novel features of retrieved information into episodic memory to a greater extent than the right prefrontal cortical regions, and (3) right prefrontal cortical regions are involved in episodic memory retrieval to a greater extent than left prefrontal cortical regions.

Following this initial proposal, and in light of new findings from fMRI experiments with better scanners and sharper experimental design, Habib, Nyberg, and Tulving (2003) reformulated HERA to be stricter and more precise in its assertions. In this reformulation there were two specific hemodynamic hypotheses that were expressed using abbreviations: 'Enc' meant encoding, 'Ret' meant retrieval, 'L' stood for a given left prefrontal cortical region, and 'R' stood for the corresponding region in the right prefrontal cortex. Combinations of task (Enc or Ret) and regions (L or R) stood for the observed activation during a given task in a given region. Thus, the two hemodynamic hypotheses that constitute the HERA model were stated: (1) (Enc L − Ret L) > (Enc R − Ret R) and (2) (Ret R − Enc R) > (Ret L − Enc L) (Habib et al. 2003, 241). Since the reformulation of HERA, several neuroimaging experiments have yielded results that supported the HERA model (e.g., Babiloni et al. 2006; Okamoto et al. 2011).

The HERA model was proposed as a description of a set of findings showing an asymmetry between encoding and retrieval tasks in episodic memory. Regardless of what large-scale theory one adopts, the HERA model stands as a set of hemodynamic findings, which came out of a data-driven approach and did not assume too much about theories of memory systems. Take away terms like semantic memory or episodic memory, and we can still talk about the HERA model in terms of specific, well-defined remembering tasks. The hemodynamic findings would still stand if theories of cognitive science change and new theories differently divide and categorize human memory, or if they even exclude categorizations. HERA is a contribution to cognitive neuroscience, and its development illustrates how experimental knowledge in neuroimaging grows. When we look at fMRI with

https://doi.org/10.1086/677691 Published online by Cambridge University Press

an eye toward appreciating the kind of knowledge it can reliably provide, it appears that a more fruitful approach is seeing neuroimaging experiments as tools for expanding our knowledge of relationships between cognition and hemodynamic processes in the brain rather than seeing them as merely novel constraints on existing theories.

Furthermore, the growth of hemodynamic knowledge can help researchers develop novel conceptualizations of cognition and its substrates in the brain. Let us recall from above that observed hemodynamic activity could be due to factors other than neuronal activity, such as the activity of glial cells. The traditional view of glia is that they passively support and maintain neurons, support neurotransmission, maintain ionic balance in extracellular space, and insulate axons to speed up action potentials. However, recent findings suggest that especially astrocytes, a type of glia, have more active roles in brain function. New research shows that glia are involved in synapse formation, modulation of synaptic function through bidirectional communication with neurons, and regulation of blood flow. Astrocytes talk to each other through waves of calcium ions, terminate neurotransmitters, and mediate their recycling. Astrocytes also talk to neurons; they monitor and respond to neuronal activity and possess the same receptors as neurons. Neurotransmitters released by neurons activate Ca-based signaling in astrocytes, which release neuroactive substances and signal back to neurons, thus forming a feedback loop and enhancing or inhibiting neuronal activity (see Bezzi and Volterra 2001; Allen and Barres 2009).

Of course, there is still a lot to learn about glia and their functions, but these new findings force a rethinking of the fMRI BOLD response with respect to the kind of activity it may point to in the brain. The traditional view of glia may have created a "neuronal bias" in cognitive neuroscience, which has come to have a neuron-centric view of the brain and cognition. Perhaps it is time to revise our neuronally inspired view and integrate the recent discoveries about glial function in our theoretical construals of cognition. When fMRI is freed from the neuronal bias, it can help greatly in such revisions as a source of hemodynamic knowledge. Since fMRI does not tell us whether increased activation is due to either glial or neuronal activity, specific inferences to glial or neuronal activity may not be reliable. However, inferences from fMRI results to increased hemodynamic activity are reliable provided that the given experiments constitute severe tests of the hemodynamic hypotheses of interest. These hypotheses can be interpreted to refer to the combined workings of glia and neurons that occur when subjects perform cognitive tasks. Thus, hemodynamic findings can be fruitful in the development of novel conceptualizations of how the human brain gives rise to cognition incorporating glial contributions to cognitive processing. These conceptualizations can usher in new connectionist or dynamic models of cognition inspired by not just neurons but also glia,

https://doi.org/10.1086/677691 Published online by Cambridge University Press

which would help researchers gain novel and broader insights into cognition and its substrates in the brain.

REFERENCES

Aktunc, M. Emrah. 2011. "Experimental Knowledge in Cognitive Neuroscience: Evidence, Errors, and Inference." PhD diss., Virginia Polytechnic Institute and State University.

———. 2014. "Tackling Duhemian Problems: An Alternative to Skepticism of Neuroimaging in Philosophy of Cognitive Science." *Review of Philosophy and Psychology.* doi:10.1007/s13164-014-0186-3.

Allen, N. J., and B. A. Barres. 2009. "Glia—More Than Just Brain Glue." *Nature* 457 (5): 675–77.

Babiloni, C., F. Vecchio, S. Cappa, P. Pasqualetti, S. Rossi, C. Miniussi, and P. M. Rossini. 2006. "Functional Frontoparietal Connectivity during Encoding and Retrieval Processes Follows HERA Model: A High-Resolution Study." *Brain Research Bulletin* 68:203–12.

Bezzi, P., and A. Volterra. 2001. "A Neuron-Glia Signalling Network in the Active Brain." *Current Opinion in Neurobiology* 11:387–94.

Habib, R., L. Nyberg, and E. Tulving. 2003. "Hemispheric Asymmetries of Memory: The HERA Model Revisited." *Trends in Cognitive Sciences* 7:241–45.

Hardcastle, V. G., and C. M. Stewart. 2002. "What Do Brain Data Really Show?" *Philosophy of Science* 69:S72–S82.

Huettel, S. A., A. W. Song, and G. McCarthy. 2008. *Functional Magnetic Resonance Imaging*. 2nd ed. Sunderland, MA: Sinauer.

Klein, Colin. 2010. "Images Are Not the Evidence in Neuroimaging." *British Journal for the Philosophy of Science* 61:265–78.

Logothetis, N. K. 2008. "What We Can Do and What We Cannot Do with fMRI." *Nature* 453:869–78.

Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

———. 2005. "Evidence as Passing Severe Tests: Highly Probable versus Highly Probed Hypotheses." In *Scientific Evidence: Philosophical Theories and Applications*, ed. Peter Achinstein, 95–127. Baltimore: Johns Hopkins University Press.

Mayo, Deborah, and Aris Spanos. 2011. "Error Statistics." In *The Handbook of Philosophy of Science*, vol. 7, *Philosophy of Statistics*, ed. Prasanta S. Bandyopadhyay and Malcolm Forster, 1–46. Amsterdam: Elsevier.

Okamoto, M., Y. Wada, Y. Yamaguchi, Y. Kyutoku, L. Clowney, A. K. Singh, and I. Dan. 2011. "Process-Specific Prefrontal Contributions to Episodic Encoding and Retrieval of Tastes: A Functional NIRS Study." *NeuroImage* 54:1578–88.

Roskies, Adina L. 2008. "Neuroimaging and Inferential Distance." *Neuroethics* 1:19–30.

———. 2010. "Neuroimaging and Inferential Distance." In *Foundational Issues in Human Brain Mapping*, ed. Stephen J. Hanson and Martin Bunzl, 195–215. Cambridge, MA: MIT Press.

Tulving, E., S. Kapur, F. I. M. Craik, M. Moscovitch, and S. Houle. 1994. "Hemispheric Encoding/Retrieval Asymmetry in Episodic Memory: Positron Emission Tomography Findings." *Proceedings of the National Academy of Sciences* 91:2016–20.

Uttal, William. 2001. *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: MIT Press.

https://doi.org/10.1086/677691 Published online by Cambridge University Press