Neuroscience and the Multiple Realization of Cognitive Functions*

Carrie Figdor†‡

Many empirically minded philosophers have used neuroscientific data to argue against the multiple realization of cognitive functions in existing biological organisms. I argue that neuroscientists themselves have proposed a biologically based concept of multiple realization as an alternative to interpreting empirical findings in terms of one-to-one structure-function mappings. I introduce this concept and its associated research framework and also how some of the main neuroscience-based arguments against multiple realization go wrong.

1. Introduction. Many nonreductive physicalists have long been united by the belief that mental properties or states are multiply realizable.[1] This consensus was established largely on the basis of the intuitions and analogies presented by Putnam (1967), Block and Fodor (1972), and Fodor

*Received October 2009; revised December 2009.

†To contact the author, please write to: Department of Philosophy, 260 English-Philosophy Building, University of Iowa, Iowa City, IA 52242; e-mail: carrie-figdor@uiowa.edu.

‡I wish to thank Jennifer Mundale, John Bickle, and the audience at the 2003 Society for Philosophy and Psychology annual meeting for comments on a distant ancestor of this paper; at least two anonymous reviewers at *Philosophy of Science*; Tom Polger and other members of the October 2008 Workshop on Multiple Realization at the University of Cincinnati, including Ken Aizawa, John Bickle, Carl Craver, Carl Gillett, Larry Shapiro, and Jacqueline Sullivan; and my colleagues in the Philosophy Department and the Program in Cognitive Neuroscience at the University of Iowa.

1. In this literature, the terms "property" and "type" are used interchangeably: e.g., mental properties are realized by physical properties, but the one-to-one relation contrasted with multiple realization is type-type identity. This mixed usage is harmless if we assume, as is usual, that particular (or token) states have properties in virtue of which they are classified by or under types, whatever the ontological status of the properties. I will follow customary usage to the extent possible to avoid drawing attention to this issue. I will use the terms "cognitive," "psychological," and "mental" interchangeably. Finally, I also adopt the common (if not universal) current usage of the term "multiple realizability" to signify possible (e.g., extraterrestrial) as well as actual (e.g., biological) cases and "multiple realization" to signify actual cases.

Philosophy of Science, 77 (July 2010) pp. 419–456. 0031-8248/2010/7703-0005$10.00
Copyright 2010 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

(1974). As many critics have recently pointed out, intuition and analogy remain a primary means to argue for multiple realization (MR) in existing creatures.[2] This is rather striking, as the scientific—particularly neuroscientific—advances that have occurred since 1967 have prompted many of those interested in the mind-body problem to wonder how this research might affect the debate.

Many philosophers who have looked at this research see bad news for MR. Bechtel and Mundale (1999), Bickle (2003), Polger (2004), Shapiro (2004), and others have used data from cognitive neuroscience, cellular and molecular neuroscience, and vision science to argue that MR in the cognitive systems of evolved biological organisms is unobvious, doubtful, implausible, or false. The cumulative effect of their arguments merits close attention, aside from its challenge to the role of intuition and analogy in the debate. First, by focusing on MR in biological creatures, the implications of the debate for clinical and experimental practices are emphasized. Second, if they are right, they will have shown that the main argument for nonreductive physicalism is scientifically unviable with regard to such creatures. The position would become as practically untenable as phlogiston theory after the discovery of oxygen. (Of course, some philosophers have long considered it untenable—e.g., P. M. Churchland 1981; P. S. Churchland 1986.) Efforts to clarify the metaphysics of realization (e.g., Gillett 2003) would also lose a good deal of their motivation.

The main aims of this article are to demonstrate the empirical viability of the MR hypothesis and to clarify the terms of an empirically based debate about its possibility in the cognitive states in existing biological organisms. To this end, I focus on research in cognitive neuroscience for two reasons. First, since neuroscience in general has been the primary source of evidence used in "empirically based" arguments against MR, this focus obeys a de facto restriction to an "evolved-biological" debate. A truly empirically based debate would also include cognitive systems we can engineer, biologically or artificially. Second, since cognitive neuroscience in particular is directly concerned with linking mind and brain in existing animals, its hypotheses, findings, and methods are prima facie relevant to a debate about the nature of that link in the cognitive systems of organisms in that class. My third aim is to rebut some prominent arguments against MR in the cognitive systems of existing animals that are also based on cognitive neuroscientific data. Since data from the same science are being used to draw the opposing conclusion from mine, it is incumbent on me to show how these arguments go astray.

2. See Shapiro (2000), Clapp (2001), and Gillett (2003); other intuitive examples are found in Kim (1993), while Keeley (2000) examines MR in electric fish.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

The article has three main sections. In section 2, I introduce the cognitive neuroscientific research program that aims to map brain structures to cognitive processes. In section 3, I motivate and explain cognitive neuroscience's version of MR. In section 4, I assess some of the cognitive neuroscientific evidence used to argue against MR in biological organisms.

2. The Cognitive Neuroscientific Background of an Empirical MR Debate. In what follows, I adopt more or less intact the coarse-grained vocabulary of "structure" and "function" that cognitive neuroscientists use when discussing the entities (properties and relations) that are among what those philosophers refer to as realizers and realizees (or realized properties). Most broadly, they are used in ways analogous to the philosopher's "physical-mental" distinction. The term "structure" is particularly protean, often being used to introduce any physical substrate, with precise terms employed later in context; to put the point in terms borrowed from the philosophical lexicon, "c-fibers" and "c-fibers firing" may both be initially called "structures." Following Bechtel and Mundale's (1999) working translation of the philosophical term "brain state" as "activity in a brain area," a working translation of not-otherwise-specified structure-function talk in cognitive neuroscience is that realizing structures are the neurological or neurophysiological properties of brain areas (networks) investigated in neuroscience, and realized functions are the cognitive capacities, functions, or dispositions (the differences among these will not matter here) investigated in cognitive psychology. The multilevel nature of neuroscientific explanation (see, e.g., Craver 2007; Aizawa and Gillett 2009) strongly implies that at least some realizer properties will be instantiated by components of entities with realized properties (e.g., properties and relations of components of a brain area may in combination realize a cognitive capacity assigned to the brain area). However, nothing in my discussion turns on whether one adopts this "dimensioned" view of realization (Gillett 2003), as opposed to the "flat" view that restricts realizer and realized properties to properties of the same individual. Nor does anything depend on my background assumption that the realization relation in cognitive neuroscience (if not universally) is properly analyzed in terms of causal role-playing (roughly, F realizes G if and only if F plays, or contributes to the playing of, the causal role that individuates G).

The dominant research program in contemporary cognitive neuroscience is localizationism, which hypothesizes that cognitive systems and brains (in particular, cerebral cortex) have parts and that particular cognitive parts are realized in or by particular brain (especially cortical) parts. This program was inspired by spectacular discoveries of specific cognitive

https://doi.org/10.1086/652964 Published online by Cambridge University Press

losses in a few patients with localized brain damage.[3] Localizationism involves more than a commitment to functional specialization. Functional specialization is compatible with more than one brain area being specialized to subserve the same function and one brain area being able to subserve more than one function. However, while functional specialization without localization has long been a recognized position (Phillips, Zeki, and Barlow 1984), researchers have often adopted assumptions (discussed below) that support inferences from mappings in which a particular brain region is solely responsible for a sole cognitive task. Localizationism is functional specializationism plus these stronger assumptions.[4]

However, it is also widely accepted that an adequate explanation of how the brain subserves cognition will require understanding neural connectivity, without which, some argue, localizationism is just the "new phrenology" (Phillips et al. 1984, 339; Uttal 2001; Friston 2002). In functional integration, researchers investigate the ways in which coactivated anatomical regions influence each other's activity.[5] Integrationists do not claim that cognitive processes are properties of the whole brain's operation or large portions of it, which is a holist position—and even holists (e.g., Uttal 2001) reject Lashley's (1929, 1950) theory of equipotentiality, in which any cortical anatomical region can in principle subserve any function. Integrationists hold that the brain can be divided into functionally specialized anatomical areas but that the activity of individual areas will not suffice to explain cognition. In effect, we can consider the dominant research program to be composed of an effort (associated with localiza-

3. These include Phineas Gage, a Vermont railroad foreman who suffered frontal lobe damage in an 1848 accident and retained much of his intelligence but whose personality changed dramatically; Broca's patient "Tan," who after damage to an area of his left hemisphere could understand language but could only utter the syllable "tan"; and H. M., whose 1953 lobotomy led to his inability to form new memories. H. M. was revealed to be Henry Gustav Molaison after his death on December 2, 2008 (*New York Times*, December 4, 2008, A1).

4. That said, in the empirical literature "localization" and "functional specialization" are often used interchangeably. Here, and in my discussion of "brain area" below, I aim to clarify concepts, not legislate usage. For similar reasons, I use "brain" and "cortical" interchangeably: although "brain structure" includes obviously distinct brain parts such as the amygdala or hippocampus, much localization research into "brain structure" seeks functionally significant divisions in cortex, which are more precisely called "cortical structures" (see also Mundale 2002; Ward 2006, 62–63).

5. Friston (1997, 21) defines functional specialization as "the expression of stereotyped patterns of neuronal activity in response to specific attributes of a stimulus, cognitive processing, or motor behaviour by specialized cortical areas, subareas or neuronal populations" and functional integration as "the interactions among specialized neuronal populations and how these interactions depend upon the sensorimotor or cognitive context."

https://doi.org/10.1086/652964 Published online by Cambridge University Press

tionism) to identify focal neural contributions to cognition and an effort (associated with integrationism) to identify systems-level models of their interaction.

Many of the empirical results of localizationism are on vivid display in the form of images of cross-sections of brains color coded to indicate areas of cortical activity that have been correlated with specific cognitive functions or processes.[6] In fMRI, the images are the fruit of a complex process that measures a net decrease in deoxyhemoglobin in a cortical area or areas. This is taken to indicate an increase in blood flow to, hence more neural activity in, that area, which in turn is taken to indicate the relevance of the area or areas for performing the cognitive task(s) being explored in the imaging experiment.[7] The broad contours of an empirically based MR debate emerge from some basic features of the structure-function mappings represented in these images.[8]

First, the images represent two distinct mapping projects: neuroanatomy and cognitive (or functional) neuroanatomy. Both projects employ the term "brain area" (or "cortical area," "structure," "region") to pick out their results but use different criteria to individuate areas. In neu-

6. Functional magnetic resonance imaging (fMRI) is popular because it is noninvasive and has relatively high spatial resolution of the cortical surface. Positron emission tomography (PET) requires injecting subjects with radioactive tracers. Other noninvasive technologies include electroencephalography, magnetoencephalography, transcranial magnetic stimulation, transcranial electrical stimulation, and diffusion (tensor or spectrum) imaging. Invasive procedures used on humans include direct cortical stimulation during neurosurgery and implantation of electrodes to locate foci of epileptic seizures; single-cell recordings and induced lesions are limited to nonhuman animals. Since different technologies have different degrees of spatial or temporal resolution, researchers increasingly use results from more than one method when developing experiments and interpreting data. See Ward (2006) for accessible explanations of technologies and Savoy (2001) for extended critical discussion of the benefits and drawbacks of each.

7. Although PET and fMRI signals are both hemodynamic, PET measures changes in blood flow while fMRI measures blood oxygenation levels (and changes in blood flow only indirectly). Both measure such changes at a scale of millions of neurons within volumes of a few cubic millimeters, with signals sampled every few minutes (PET) or seconds (fMRI) from many (e.g., 100,000) cortical positions (voxels). Note that much cognitive processing, which occurs at speeds of milliseconds, is invisible even at sampling rates of every few seconds. Attwell and Iadecola (2002) suggest the fMRI signal reflects postsynaptic neurotransmitter activity, not (as usually assumed) energy use in presynaptic terminals or glia.

8. Throughout this article, the term "structure-function mapping" is intended to be neutral regarding the differences in the direction of inference (function to structure or structure to function) that leads to a given mapping in a particular experimental paradigm. In the empirical literature in general, reference to "one-to-many" and "one-to-one" mappings are often and easily disambiguated in context, although only the former phrase threatens any genuine confusion.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

roanatomy, brain areas are individuated by distinctions in cellular mechanisms, cytoarchitecture, morphology, myelination, axonal projections and connectivity, and neurophysiology, without essential reference to an area's possible role in supporting cognition, even if such a role motivates the mapping effort. For example, Korbinian Brodmann's (1909/1914) map of 47 human cortical areas, still used as a reference point in both mapping projects, was based on purely anatomical criteria (differences in cell types and their distribution, or cytoarchitecture). Similarly, Hagmann et al. (2008) propose a neuroanatomical map of major axonal connections in human cortex on the basis of anatomical data from diffusion imaging, which traces the diffusion of water through brain tissue to reveal the orientation of axon fibers. Neuroanatomical maps contain what I will call "anatomical areas," which have names like "inferotemporal cortex" or "Brodmann's area 44."

In cognitive neuroanatomy, brain areas are individuated using anatomical and cognitive-functional criteria. The resulting areas are actually structure-function mappings, and cognitive-neuroanatomical area names pick out these mappings. For example, the functions of "early" vision—edge, orientation, contrast and brightness detection—are not mapped to V1; they are mapped to the medial calcarine sulcus in humans, and this mapping is called V1. Similarly, a pioneering cognitive neuroanatomical study by Felleman and Van Essen (1991b) distinguished 32 visual areas in macaque cortex on the basis of connectivity, architectonics, topographic organization of the visual field (retinotopy), distinct receptive fields of neurons, lesion and stimulation studies, and prior studies that distinguished visual cortical areas using similarly varied criteria. This suite of criteria is widely used in vision research, and similar multiple criteria are used for individuating neurocognitive areas in general.[9] Cognitive neuroanatomical maps contain what I will call "cognitive areas," which have names like "V1" or "Broca's area."[10]

Second, hypothesized structure-function mappings fix the reference, not the meaning, of cognitive-area names. That is why it is not a conceptual

9. Major individuative criteria for visual areas include (1) cyto- and myeloarchitecture, (2) connectivity, (3) retinotopic organization, and (4) cognitive function, as revealed by single-cell recordings, lesion studies, and neuroimaging analyses (Orban, Van Essen, and Vanduffel 2004; see also Felleman and Van Essen 1991a, 5). Retinotopic organization (retinotopy) refers to the layout of cortical neurons processing visual information relative to the layout of the input on retinal cells (e.g., points close in space on the retina are close in space in V1). Keeley's (2002) multicriterial analysis for individuating the senses may be considered a special case of cognitive-area individuation methods.

10. Henson (2005) distinguishes "areas" (anatomical divisions, e.g., Brodmann's) from "regions" (functionally significant anatomical divisions); Phillips et al. (1984) use "cortical area" as I use "cognitive area."

https://doi.org/10.1086/652964 Published online by Cambridge University Press

truth that vision is the function of visual cortex or conceptually incoherent for auditory information to be processed in visual cortex (Von Melchner, Pallas, and Sur 2000; Burton 2003). Either or both elements in a hypothesized cognitive area may be modified following further research. For example, Broca (1861/1960/2001) hypothesized that speech production was located at the superior temporal gyrus at approximately Brodmann's areas 44 and 45. This mapping is called Broca's area. However, speech production is now thought to be in more distributed prefrontal areas, and Broca's area also appears to play a role in processing natural language syntax, musical syntax, perception of rhythmic motion, imaging movement trajectories, and conducting local visuospatial searches (Marshall and Fink 2003; Grodzinsky and Santi 2008). Such modifications of named structure-function mappings exemplify difficult problems in the theory of reference (Field 1973)—in particular, the issue of when a term refers to the same entity that has changed over time or whether it has changed its reference. However, since an empirical MR debate at this time involves examining the results of rapidly developing disciplines, the possibility of at least some referential indeterminacy must be expected and tolerated by both sides.

Third, functions may be mapped to anatomical networks as well as areas, but the term "network" is also ambiguous (Henson 2005, 215–16). A weak cognitive network is a set of coactivated anatomical or cognitive areas that subserve a cognitive function. A difference in component areas suffices for a difference in a weak network.[11] A strong cognitive network is a set of coactivated anatomical or cognitive areas that exhibit effective connectivity, or activity- and time-dependent influences on processing between areas. A difference in the type of interaction between component areas suffices for a difference in strong networks. Networks of either type can share components, but the contribution of an area to subserving a function may differ when it participates in different networks.[12] Also,

11. Thus, Kosslyn (1999, 1284) notes that it is often misleading when researchers talk of brain circuits: "in most studies all that is revealed are a set of activated (and/or deactivated) areas, with no information about the flow of information between the areas. Thus, what we are seeing are the footprints of components of the functional architecture that are evoked during the task, but we do not see a specific circuit."

12. For example, suppose each of several areas makes a simple cognitive contribution to the performance of object recognition (Kosslyn et al. 1994). It does not follow that each area is unipotential, or specialized to perform one simple function. Activity in the component areas may shift in response to different conditions of processing or to temporary or permanent damage, and there may be considerable idiosyncrasy in individuals or across species regarding which areas realize which simple functions. In such cases, simple functions could be realized in more than one area within and across species, even though the function they subserve as a whole remains the same. These cases provide possibilities for degeneracy, discussed below.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

component areas may not individually suffice for a cognitive function, but if at least one does—that is, if a component is a cognitive area and not just an anatomical area—any cognitive network in which it participates would constitute an additional cognitive-area layer. In principle there may be many such layers.

Outside of extensively studied peripheral cognitive areas such as visual, sensorimotor, auditory, and motor cortex, few results of localizationism to date are entirely uncontroversial. Researchers are acutely aware that to implicate an anatomical region in the performance of a function is not to localize that function in that region. Cognitive neuroanatomy is thus fertile ground as the basis of an empirical MR debate. As I will argue below, however, such a debate is not adequately framed merely by restricting the relevant evidence to that gleaned from cognitive neuroscience. Inter alia, it requires an appropriately stated MR hypothesis. As it happens, biology provides us with one.

3. MR as Degeneracy in Cognitive Neuroanatomy. The need for an MR hypothesis suited to an empirically based debate is motivated by the inadequacy in that context of the various theses regarding multiple realizability that can be derived from the philosophical literature. These include:

> **Weak MR**. At least some creatures that are not exactly like us in their physical composition can be conscious.
> **SETI (search for extraterrestrial intelligence) MR**. Some creatures that are significantly different from us in their physical composition can be conscious.
> **Standard MR**. Systems of indefinitely (perhaps infinitely) many physical compositions can be conscious.
> **Radical MR**. Any (every) suitably organized system, regardless of its physical composition, can be conscious (Polger 2004, 6).[13]

None of these hypotheses is appropriate in a debate about MR in existing biological organisms. Implicitly, all reflect their origin in an intuition-based debate about multiple realizability that included extraterrestrial beings, computers, advanced robots, and other hypothetical possibilities as well as evolved biological animals. MR, like its main rival the identity theory, is fundamentally a claim about the relation between mental states and their physical substrates. It is not about the scope of this relation (even when restricted to evolved biological creatures). So *pace* Weak,

13. Polger focuses on consciousness, but these theses can be suitably reformulated for any mental state. Note that Polger uses "MR" for multiple realizability, not multiple realization (see n. 1).

https://doi.org/10.1086/652964 Published online by Cambridge University Press

SETI, and Standard MR, it is not essentially about the number of different kinds of minded creatures or the potential number of different physical substrates for which this relation may hold. The issue of differences between creatures is independent of the issue of differences between physical substrates, and, empirically speaking, MR may turn out to be true within one biological species only. Nor do qualitative differences (between creatures or cognitive systems) have a place in an empirical debate: "not exactly like" and "significantly different" are not subject to empirical tests unless quantified, and "indefinitely (perhaps infinitely) many" is not subject to empirical tests at all. Radical MR suffers somewhat from the qualitative nature of what counts as "suitably" organized, but its main problem is that it is consistent with the possibility of one suitably organized substrate for each mental state and so cannot capture what is essential to MR. Finally, all four theses depend on an empirically irrelevant distinction between composition and structure (in the intended sense of "arrangement") in the individuation of physical realizers. As noted above, anatomical areas are individuated on the basis of multiple criteria that include both: cytoarchitecture is a type of composition; connectivity is a type of structure.[14]

Instead, the relevant sciences can provide us with an appropriately stated hypothesis. In biology, degeneracy is defined as the ability of elements that are structurally different to perform the same function or yield the same output (Edelman and Gally 2001).[15] Thus broadly defined, degeneracy is found at all levels of biological organization, from the molecular, cellular, and genetic levels up to the level of organism: different nucleotide sequences encoding the same polypeptide, different antibodies binding the same antigen, different patterns of muscular contraction yielding the same movement, and different encodings of the same message. As a biological hypothesis, degeneracy has been posited to explain a number of studies of biological organisms, from yeast to humans, in which striking structural differences at various suborganism levels appear to have little or no organism-level effects.[16] Edelman and Gally argue that degeneracy

14. I return to the issue of which physical differences matter for MR below.

15. Tononi, Sporns, and Edelman (1999, 3257) note that the term "degeneracy" is taken from immunology, where it refers to the ability of different antibodies to bind to the same antigen.

16. See Edelman and Gally (2001) and references therein for detail on the following studies. In "knock-out" mice, in up to 30% of cases there is little or no phenotypic difference in mice that lack the genes to produce myoglobin, tenascin C, vimentin, and other important proteins. In yeast, systematic screening of single-gene deletions at more than 500 gene loci shows that fewer than half of the cultures had any quantitative growth defects. In *Drosophila*, when either the gene for fasciclin (a cell-adhesion protein on the surface of *Drosophila* neurons) or the gene for the cytoplasmic Abelson tyrosine kinase

https://doi.org/10.1086/652964 Published online by Cambridge University Press

is a prerequisite of natural selection, which requires genetic dissimilarity in a population to operate but must avoid the likely lethality of most mutations if individual genes are wholly and uniquely responsible for phenotypic traits. At the genetic level, a natural solution is overlapping networks of unrelated genes that can, given appropriate conditions for gene expression, produce the same outcome.

The biological concept of degeneracy has been appropriated and developed by cognitive neuroanatomists as an alternative hypothesis to one-to-one mappings (Price and Friston 2002; Friston and Price 2003; Noppeney, Friston, and Price 2004). In cognitive neuroanatomy, degeneracy is the claim that, for a given cognitive function F, there is more than one nonisomorphic (nonidentical) structural element that can subserve F, either within an individual at a time, across individuals, or within an individual across times.[17] In an empirically based MR debate centered on neuroscience, MR is just degeneracy in cognitive neuroanatomy, or, more precisely, since there are cases of degeneracy that do not or may not count as MRs, MRs are special, perhaps paradigmatic, cases of degeneracy. In the rest of this section, I will introduce this concept as it is used in cognitive neuroscience, before turning to issues in the metaphysics of realization and MR that bear on this conceptual assimilation.

Direct empirical motivation for hypothesizing degeneracy in cognitive neuroanatomy stems in part from long-standing anomalies for localizationism based on lesion studies. These problems include differences in deficits with similarly located lesions (and vice versa) and restitution of function after damage (a form of plasticity). Such anomalies are what led Lashley to propose his theories of equipotentiality and mass action

is completely deleted, there are no gross abnormalities in nervous system development, even though these proteins have no obvious structural or functional similarity, but major defects result when both are deleted. In humans, subjects who had exhibited no psychological abnormalities were found in fMRI scans to lack the corpus callosum that normally connects the two cerebral hemispheres (although subsequent detailed psychological testing did uncover subtle abnormalities in their functioning).

17. I take this to include degeneracy in comparative cognitive neuroanatomy (i.e., across species): degenerate functions may be unique to humans (e.g., reading), human but not uniquely so (e.g., motion detection), or (in principle) nonhuman (e.g., echolocation). The researchers cited in the text do not mention cross-species possibilities, presumably since the goal of their research, like that of localizationism in general, is to explain human cognition. (I discuss animal models in sec. 4.) Also, if degeneracy's requirement of more than one structure seems weak—e.g., compared to the demand for indefinitely (perhaps infinitely) many realizers in Standard MR—recall that Kim's (1993) influential discussion of jadeite/nephrite has long been accepted as sufficient both to characterize multiple realizability and to raise the theoretical issues that have dominated the philosophical debate and threatened the nonreductive physicalism that multiple realizability (or realization) is used to support (e.g., Fodor 2000).

https://doi.org/10.1086/652964 Published online by Cambridge University Press

(whereby lesion size, not location, determined the deficit) in the first place.[18]

A major motivation, however, has been anomalous results from the imaging studies that have come to dominate cognitive neuroscientific research in recent decades.[19] These studies frequently show areas of activation that differ across subjects (or within a subject at different times) within the same experimental paradigm (see fig. 1). These differences, which are usually ignored as "noise" or "random error," are not simply a reflection of the inevitable differences between individual brains in the precise location of gyri and sulci. Such anatomical idiosyncrasies are usually standardized by fitting individual imaging results to a common brain template (e.g., the Talairach and Tournoux [1988] atlas).[20] In some cases, the imaging results replicate the earlier anomalies based on lesion data.

18. Phillips et al. (1984, 328–33) also trace Goltz's resistance to localization to the restitution of function and generalized (rather than specific) impairments after lesions. (Goltz used dogs in his lesion studies.)

19. Lesion and imaging studies are broadly complementary. In imaging, we manipulate function in order to infer (ideally) that certain cortical structures are sufficient for a function. In lesion studies, we manipulate cortical or subcortical structure (or it gets "manipulated" by stroke, etc.) in order to infer (ideally) that certain structures are necessary for a function. However, even in the best of cases (e.g., no intersubject or intertrial variation in imaging studies, no difference in functional deficit in lesion studies), many alternative interpretations of the results of manipulation are available. For example, in fMRI, the blood-oxygen-level-dependent signal does not distinguish excitation from inhibition (Ward 2006) or neural codes that involve timing and synchronization. Thus, images can indicate areas that are active not because of what they are doing but because they are being prevented from doing something or are trying to do something but not succeeding or simply because a task requires more of a general resource (e.g., attention). Inference to the role of a structure from lesion studies is also difficult; to borrow Gregory's (1961) famous analogy: if removing a resistor from a radio causes it to emit strange howls, it would be incorrect to ascribe to the resistor the function of howl suppression. There may be no deficit if the lesioned area's function is subserved by undamaged systems that modify their operations or by newly created components (these are cases of degeneracy, explained in the text). Also, a lesioned area may facilitate performance without being necessary for it. Other concerns include whether data from single cases or groups should be used in lesion studies (Caramazza 1986) and whether it is legitimate to "standardize" brains (e.g., using the Talairach and Tournoux [1988] brain template) or average imaging results across individuals (Savoy 2001). Of course, not all localization claims are unreliable; e.g., we know the hippocampus plays an essential role in memory formation, and the occipital cortex processes visual information. The point is that many inferences to structure-function mappings (in either direction) from lesion or imaging data are not highly confirmed at this time.

20. The idiosyncrasies can be significant. On the basis of a survey of abstracts for the 2000 International Conference on the Functional Mapping of the Human Brain, Savoy (2001, 26) notes that irregularities of cortical size, shape, and foldings across subjects are a serious concern for cross-subject image comparisons.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

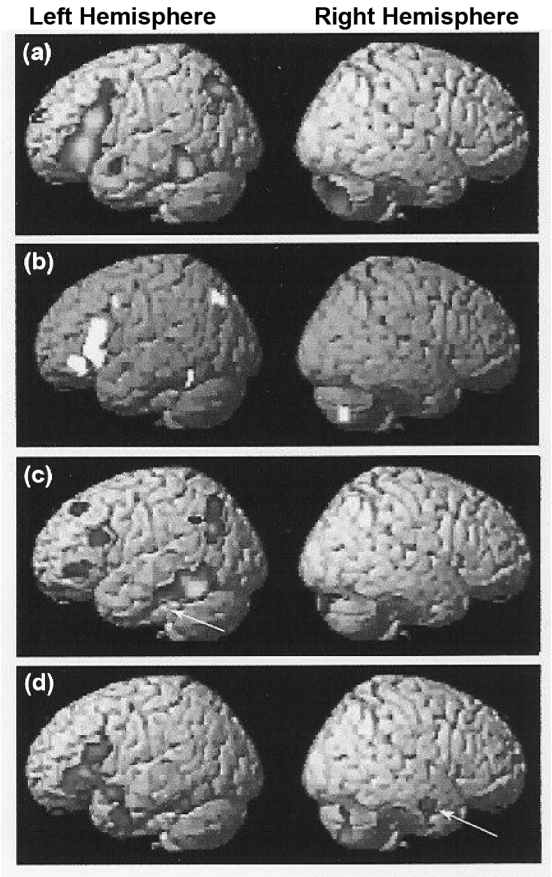


Figure 1. Normals based on previously published studies involving a semantic task (e.g., matching words and pictures). *a*, Results averaged over 12 subjects, without distinguishing between areas found in one subject or more than one; *b*, results common to all 12 subjects only; *c*, data from subject 1 only; *d*, data from subject 2 only. Arrows indicate areas of activation in individuals that do not appear in averaged images (*a* and *b*). Source: Price and Friston (2002). Color version available as an online enhancement.

For example, double dissociations have been found between Broca's aphasics and neural activity in Broca's area and between Wernicke's aphasics and neural activity in Wernicke's area—that is, there are patients with lesions in Broca's area who do not have Broca's aphasia and patients with Broca's aphasia who do not have lesions in Broca's area (ditto,

https://doi.org/10.1086/652964 Published online by Cambridge University Press

mutatis mutandis, for Wernicke's aphasia and area; Dronkers, Redfern, and Knight 2000).[21]

In other cases, new anomalies arise from combining imaging and lesion data. Price and Friston (2002) examined results from fMRI studies of normal subjects performing semantic-processing tasks (e.g., picture naming) and fMRI studies of lesion patients capable of performing the same tasks at normal levels. The patients' lesions were plotted to the same standard brain plan used for normal subjects. Although the lesions were located in the areas activated in the normal subjects' performance of the tasks, the fMRI data from the patients showed entirely distinct areas activated during their performance (see fig. 2). They conclude that none of the cortical areas activated in the normals are in fact necessary for the tasks.

The degeneracy hypothesis can explain these puzzling processing differences and data showing undeniable cross-subject specialization of function in cortex that have made Lashley's hypotheses untenable. Complementary to degeneracy is the concept of pluripotentiality: when a single structure subserves more than one function. Pluripotential structures fill a spectrum of degrees of functional specialization between unipotentiality and equipotentiality. There can be degeneracy without pluripotentiality (if two unipotential structures subserve the same function) and pluripotentiality without degeneracy. But degeneracy and pluripotentiality are considered closely associated for evolutionary reasons. Degenerate structures will typically exhibit the variability that selection processes require and so will not be duplicates. As a result, many degenerate structures will be such that they can produce the same output in one processing context but different outputs in another. For example, a given structure may subserve function F when activated within one network and function G when activated within another, or if there are two structures that normally subserve F and G, respectively, and the G-structure is damaged or inhibited, the F-structure must be pluripotential if it can step in to subserve G. Thus, for any cognitive function F, degeneracy can occur if there is (i) more than one unipotential nonduplicate area (or network), each suf-

21. Classic double dissociations are cases in neuropsychology when one (lesioned) subject or group (A) demonstrates one type of cognitive loss or impairment (F_1) while another capacity (F_2) is left intact or relatively so, and a second subject or group (B) demonstrates the loss or impairment of F_2 while F_1 is intact or relatively so. A classic case is phonological dyslexia and surface dyslexia: subjects with surface dyslexia have trouble reading irregular words ("chef") but no (or less) difficulty reading nonwords ("mave"), while subjects with phonological dyslexia exhibit the opposite difficulty. The concept has been extended to include cognitive areas, even though there is significant controversy as to what legitimately can be inferred about cognitive organization from double dissociations (Shallice 1988; Farah 1994; Plaut 1995; Coltheart and Davies 2003; Dunn and Kirsner 2003).

https://doi.org/10.1086/652964 Published online by Cambridge University Press

https://doi.org/10.1086/652964 Published online by Cambridge University Press

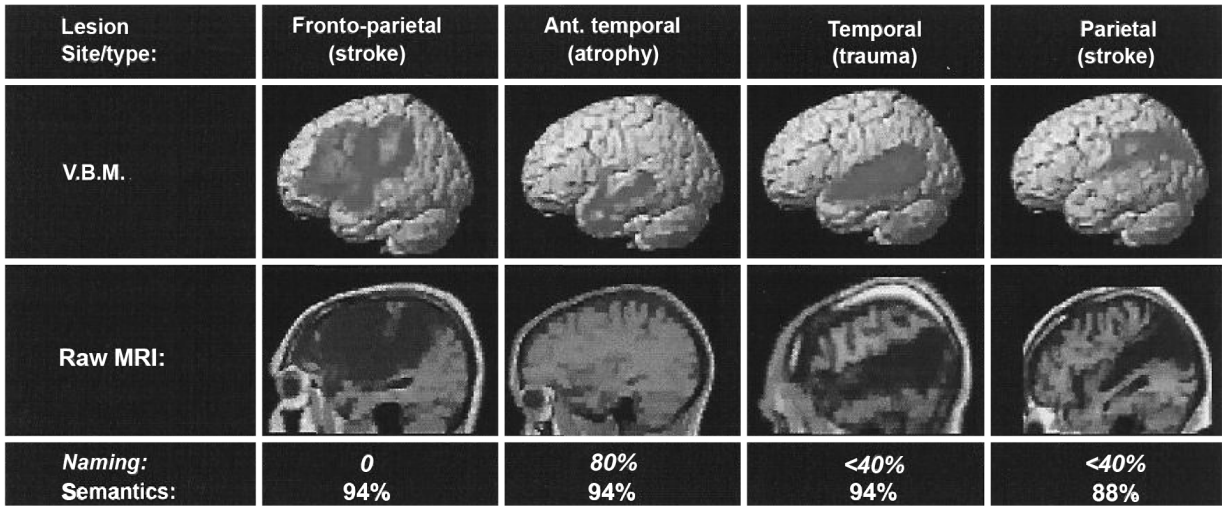


Figure 2. *Top row*, site, type, and extent of lesions (measured in voxel-based morphology) in areas activated in normals during the semantic paradigm used for the images in figure 1. Percentages show performance by the patients on these semantic tasks. None of the areas activated in normals appear necessary. Source: Price and Friston (2002, 419). Color version available as an online enhancement.

ficient for F; (ii) more than one pluripotential nonduplicate area (or network), each of which suffices for F in certain contexts; or (iii) more than one combination of unipotential or pluripotential nonduplicate areas (or networks) that in combination suffice for F. This last possibility includes degenerate networks that partially overlap in their component areas. Overlapping networks may be considered cases of only partial MR if each common area makes the same contribution to each network but not if their contributions differ between networks.

A measure of the degree of degeneracy of a particular structure-function relationship is given by the number of sufficient wholly disjoint elements (areas or networks) that can produce the same output. This number—the order of degeneracy—can be determined behaviorally by the minimum number of anatomical areas that must be lesioned before a behavioral deficit can be observed.[22] The order number is not the same as the number of sufficient systems that may subserve the same function since lesioning one area common to more than one partially overlapping degenerate network can result in a cognitive deficit and would still be a case of first-order degeneracy. In second-order degeneracy, there are at least two wholly disjoint structures subserving the same function. In short, the same order number (the same degree of degeneracy) may reflect different numbers of realizations.

Within this general framework, not all cases of degeneracy do or may count as MRs. For example, when degenerate systems that perform the same function are coactivated, they are said to be functioning redundantly (i.e., inefficiently), even if they perform the function in distinct ways. Since degenerate systems can be latent, such that only the prepotent system operates unless deactivated, redundant functioning entails degeneracy but not vice versa. However, duplicate anatomical areas subserving the same function, whether or not they function redundantly, would no more count as MRs than the kidneys, which are both anatomically and functionally redundant.

However, further differences may stem from the selection of the relevant functions and structural elements for mapping. Degeneracy, like MR, is relative to the levels of psychological function and biological organization in a given mapping. It is an open question whether there is a right or optimal level (or range) for mapping functions and structures, even within

22. Note that this measure can only determine the order of degeneracy of a mapping within an individual.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

the limits of cognitive neuroscience.[23] In cognitive neuroanatomy, the levels of function are implicitly determined by the task analysis and the psychological and neurophysiological measurements being employed in a given study. Currently, behavioral or behavior-based measures—for example, direct responses to perceptual stimuli, performance on standard neuropsychological test batteries, double dissociations—are used to decompose and characterize functions.[24] Such measures have yielded a level of cognitive "strategies" ("routes," "pathways") in addition to functions or outputs because of the discovery or hypothesizing of more than one way of generating the same output. Outstanding among dual-route models is Ungerleider and Mishkin's (1982) discovery of two cortical routes for visual processing after initial processing in visual cortex. Similarly, reading may be performed via a lexico-semantic route or an orthographic-phonological route. However, while dual-route models are considered cases of degeneracy, they may not be MRs if cognitive functions are individuated by routes (e.g., reading by the semantic route vs. reading by the phonological route). A conservative policy would hold that there is MR only if, for any strategy that yields F, the same strategy (or the same step in at least one strategy) is subserved by different structures.[25]

23. Price and Friston (2002, 418) and Noppeney et al. (2004, 434–35) recognize that degeneracy is sensitive to the levels of function and structure. The fundamental problem is that MR (or the identity theory, for that matter) becomes trivial if psychological functions at any level of abstraction can be mapped to physical substrates at any level of abstraction in realization-relevant mappings. Henson (2005, 217–19), also noting the need to identify "the appropriate level of functional/structural abstraction" for mappings, suggests defining the appropriate level as that at which one assumes a priori that there is a one-to-one mapping; experiments would then test whether this hypothesis is correct (see also n. 32 below). This proposal has the merit of voiding Bechtel and Mundale's (1999) charge that philosophers made MR seem plausible by mapping coarse-grained psychological types to fine-grained neural types based on their intuitions.

24. This dependence on behavioral studies for labeling cognitive processes has been challenged by Price and Friston (2005). However, they suggest augmenting such data, not eliminating them, and explicitly leave open the possibility of degeneracy, even in an improved naming system. My point here is to illustrate one of the ways in which degeneracy and MR can come apart.

25. This may well be too conservative. Shapiro (2000, 644; 2004) has argued that what matters for realizer individuation are differences in causally relevant (or R) properties—"in properties that make a difference to how they contribute to the capacity under investigation." Whether or not this criterion rules out many intuition-based examples of MR, it implies that reading is multiply realized if the dual-route model is correct since the causally relevant properties used in cognitive neuroscientific individuation include connectivity, and connectivity between two routes obviously differs. Dual-strategy models have also been proposed for (e.g.) object constancy and face constancy (Ward 2006, 110), mental rotation (Kosslyn et al. 1998), and verbal response selection (Raichle et al. 1994). The terms "process," "strategy," and "route" are often used interchangeably (and may be referred to as "functions").

https://doi.org/10.1086/652964 Published online by Cambridge University Press

On the neuroanatomical side, single cells, neuronal populations, anatomical areas, or anatomical networks are among the "structural elements" that may appear in degenerate mappings. This list also may be too inclusive for MR. For example, Edelman and Gally (2001, 13765) consider cognitive neuroanatomy very likely to be highly degenerate, but they appear to individuate neural structures such that a single difference in connectivity suffices for a distinct structure.[26] But presumably MR could occur within an individual area if it contains multiple functionally specialized neuronal populations that can perform the same function or if cells within the same population switch between different encodings ("remap") at short timescales (Johnson et al. 2009). A conservative policy would count only properties of anatomical areas and networks as (possibly degenerate) realizers.

However, even a focus on anatomical areas or networks, or more generally to structural units well above molecular mechanisms, is neither arbitrary nor restrictive. Obviously, if specific cognitive phenomena are produced neither by the whole brain's operation nor by individual neurons, there must be intermediate units to which functions can be mapped. Anatomical areas fit this basic requirement. More important, the multiple criteria used to individuate anatomical areas are drawn from different levels of neuroscience, not by cortical analogues of latitude and longitude (see n. 9 and associated text). Even if cellular and molecular neuroscience are the central disciplines within or of neuroscience (Bickle 2003), cognitive neuroscience incorporates the lower-level findings: mechanisms are already among the criteria for individuating anatomical areas, and single areas can contain multiple functionally specialized units. This is why it is at best an unfounded assumption to think of anatomical areas as a mere structural stopgap until we find out more about (e.g.) mechanisms, such that these lower-level discoveries will settle the debate.[27] There is no

26. In this light of Edelman and Gally's (2001) very fine-grained individuation, it is ironic that Bechtel and Mundale (1999, 178) diagnose the apparent plausibility of MR as the result of a "methodological error" by philosophers through "mismatching" coarsely individuated functions with finely individuated realizers based on their intuitions; they conclude that "when a common grain size is insisted on, as it is in scientific practice, the plausibility of multiple realizability evaporates."

27. Bickle (2003) employs detailed evidence from cellular and molecular neuroscience mainly to make a case for reduction. His direct attack against MR (131–61) rests on empirical findings that the consolidation of memory-like capacities (including sensitization—the heightened responsiveness to noxious stimuli by an organism's defenses—and classical conditioning) in the sea slug (*Aplysia californica*) and fruit flies (*Drosophila melanogaster*) is controlled by the same cellular and molecular mechanisms, plus the claim that evolution conserves molecular mechanisms. This may all be true, but it does not show that MR is false unless (inter alia) these mechanisms are exclusive and play the same role in higher animals. Even if evolution conserves molecular mechanisms, if Ed-

https://doi.org/10.1086/652964 Published online by Cambridge University Press

empirical reason to think the list of individuative criteria of anatomical areas is closed or that it will be pruned to a single (lower-level) criterion, effectively eliminating the relevance of anatomical areas in structure-function mappings. Neuroscientific practice suggests that future discoveries will lead to a progressively more articulated taxonomy of anatomical areas, not abandonment of these structural units. The increasingly fine-grained individuation of cognitive areas within visual cortex illustrates and foreshadows this methodology.[28] There is also no empirical reason why distinctions at even lower levels, for example, in basic metabolic processes, might not play a role (*pace* Bechtel 2006, 498). If metabolic differences (or subatomic particle differences, for that matter) can be manipulated to make a measurable cognitive-functional difference in the relevant behavioral tests, the most likely result is that such processes will be added to the list of individuative criteria and weighed against the other items when conflicts arise.

Thus, while it is likely that degeneracy and MR in the cognitive neuroscientific context do not exactly coincide, I have largely left open the extent to which they may come apart. But trying to pry them apart now would essentially require using individuation criteria that either are not employed in the relevant sciences or, if they are, are among multiple individuation criteria that span biological levels. Multiple criteria imply that it is an open empirical possibility that when two neural structures count as the same by one criterion, they may count as different based on other criteria. Which of these physical differences will triumph in individuation in particular cases of criterial conflict cannot be determined a priori. (In sec. 4, I discuss a case of conflict of this sort.) In short, at least

elman and Gally (2001) are right, it will conserve degenerate molecular mechanisms. Moreover, Bickle assumes the behavior of these relatively simple creatures essentially requires psychological explanation—a claim that is controversial, even for chimpanzees and other higher primates (Andrews 2008). So even if sea slugs and fruit flies have a mechanism that explains simple behaviors that are broadly analogous to what creatures with memory may do in broadly analogous situations, and even if we also have this mechanism, nothing follows immediately about MR.

28. Savoy (2001, 10–12), displaying a 1957 map of cognitive areas in human cortex based on data from lesion studies and cortical stimulation during neurosurgery, notes that the map is "remarkably accurate" and that new technologies have enabled us to "refine" the map and add information about subcortical structures. Chemistry provides an independent scientific precedent (Le Poidevin 2000). As new behavioral differences between molecules were discovered, chemists individuated chemical kinds progressively more finely by adding criteria—moving from individuation by proportions of each kind of atom (expressed in the molecular formula) to individuation by proportions and arrangement and kinds of bonds (expressed in the structural formula) to individuation by proportions, arrangement, bonds, and orientation in space (expressed in conventional notations representing three-dimensional arrangements of atomic groups).

https://doi.org/10.1086/652964 Published online by Cambridge University Press

some degenerate systems are very likely to count as cases of MR, however the issue of realizer individuation is settled.

4. Cognitive Neuroscience and Arguments against MR. The preceding sketch of degeneracy in cognitive neuroanatomy suffices to show that MR is a viable empirical hypothesis within cognitive neuroscience. In this section, I respond to arguments based on data from cognitive neuroscience that conclude that MR is empirically implausible. First, I show how a popular strategy for using empirical data in the MR debate goes awry. This strategy involves drawing an implication from MR regarding the autonomy (in some sense) of psychology from neuroscience and then using neuroscientific data to show that this implication is false. Both Bechtel and Mundale (1999) and Shapiro (2004) develop ANA (Arguments from Nonautonomy, as I call them), although for space reasons I can only discuss Bechtel and Mundale's version here.[29] I then turn to a second argument by Bechtel and Mundale that cognitive neuroscientific methodologies show that MR is implausible. I show that they do not.

Bechtel and Mundale (1999) defend the following two claims, each of which plays a central role in the empirically based arguments against MR just sketched:

> **Claim 1**. Neuroscientific information has been useful in psychology (e.g., in guiding the decomposition and understanding of cognitive systems).[30]
> **Claim 2**. Cognitive researchers assume (implicitly) that MR is actually false.[31]

I will be arguing that the truth of claim 1 does no harm to MR and that claim 2 is false.

Claim 1 reflects Bechtel and Mundale's (1999) main concern with the

29. Given his interpretations of MR and autonomy, Shapiro's version (ANA^{Shap}) may be stated as follows: (1) MR claims that humanlike minds can be realized in very many humanlike or nonhuman-like brains that might have evolved on earth consistently with physical law. (2) If MR is true, we should be able to infer little or nothing about human brains from facts about human psychology. (3) But we can infer facts about human brains from facts about human psychology. (4) So MR is "much less obvious than philosophers suppose" (2004, xiii).

30. "We have tried (i) to demonstrate that the claim that psychological states are multiple [*sic*] realized has not been demonstrated, at least within animal life forms, (ii) to show how denying MR allows fruitful use of neuroscience in guiding the decomposition and understanding of cognitive systems, and (iii) to diagnose why multiple realizability has been so widely accepted" (Bechtel and Mundale 1999, 204).

31. I provide textual evidence of this claim below.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

"common corollary" of MR regarding autonomy, which appears as premise 2 in their Argument from Nonautonomy ($ANA^{B\&M}$):

1. MR claims that the same psychological state (process) can be realized by different brain states.[32]
2. If MR is true, information about the brain should be of little or no relevance to understanding psychological processes (cognitive systems).[33]
3. But neuroscientific data have been useful in guiding the decomposition and understanding of cognitive systems (psychological processes; claim 1).
4. So MR is empirically implausible (or false).

This argument is either unsound or invalid depending on how premise 1 is interpreted. If premise 1 is understood as an expression of the hypothesis of degeneracy in cognitive neuroanatomy, it does not imply the consequent in premise 2. Degeneracy is cognitive neuroscience's yes answer to the question of whether more than one anatomical structure may subserve the same cognitive function. It cannot rule out the use of whatever information cognitive neuroscientists want to use to determine whether that answer is correct or imply that a hybrid subdiscipline like cognitive neuroanatomy should not be possible. So on this interpretation, premise 3 is perfectly compatible with premise 1. (The implicit assumption in the consequent of premise 2 that psychologists should not find neuroscientific data useful is discussed below.)

If premise 1 is interpreted as any one of the intuition-based MR theses stated in section 3 (which seems to be Bechtel and Mundale's intent), the argument involves an equivocation.[34] Intuition-based MR (in any version)

32. "The claim of multiple realizability is the claim that the same psychological state can be realized by different brain states. Thus, it is claimed that there is a many-to-one mapping from brain states to psychological states" (Bechtel and Mundale 1999, 176).

33. "One common corollary of this rejection of the identity thesis [i.e., MR] is the contention that information about the brain is of little or no relevance to understanding psychological processes" (Bechtel and Mundale 1999, 176).

34. "The guiding assumption [in artificial intelligence] was that if mental activities could be characterized in terms of operations in a system, then they should be able to be implemented in different hardware or wetware, thereby providing alternative realizations. Taking this a step further, many philosophers became convinced that the same mental activities could be realized in brains of aliens with radically different composition from ours. The upshot of these speculations about artificial and alien minds is a metaphysical claim that mental processes are the operations themselves, and are not identified with whatever biological or other substances realize them. For the most part, we will have nothing to say about these speculative arguments, nor are we primarily concerned with the metaphysical claim. Our primary concern, rather, is with the im-

https://doi.org/10.1086/652964 Published online by Cambridge University Press

and its "common corollary" (assuming that premise 2 correctly states that corollary) were formulated in a context that included alien and artificial cognitive systems—that is, "silicon-based extraterrestrials, computers, androids, robots, and other brainless science fictional beings," as Bickle (1998, 114) puts it.[35] So if premises 1 and 2 are interpreted in their original forms, they are true (if true at all) of the cognitive systems of these beings as well as those of existing biological organisms. But premise 3 is only about the cognitive systems of the latter. In effect, the consequents of premise 2 and premise 3 are related as genus to species; put otherwise, the term "brain" in premises 1 and 2 means roughly "physical substrate," while in premise 3 "cognitive systems" means "evolved-biological brain." Neuroscientific information may be helpful in understanding biological-brain-based cognitive systems, but this is compatible with the claim that this information will not be helpful in understanding cognitive systems in general. So on this interpretation, all the premises can be true, but the conclusion does not follow.

It is worth noting that the empirical studies Bechtel and Mundale (1999) use to support premise 3 do not actually do so. These include Ungerleider and Mishkin's (1982) influential research showing two distinct neural pathways for post-V1 visual information processing in the macaque. This research led to the discovery of similar neural routes in post-V1 human vision. The error is to interpret this research as showing that the discovery of two cortical pathways (neuroscience) suggested that vision splits into two functional streams (psychology). Ungerleider and Mishkin used psychological (visual) as well as neural information to obtain their cognitive neuroanatomical results. In other words, their research (and the other studies Bechtel and Mundale cite, involving cell-level research in early visual areas) is of a hybrid nature from the start. Otherwise they would have had no reason to distinguish (or even ability to identify) these two

---

plication drawn from the multiple realizability argument that information about the brain is of little or no relevance to understanding psychological processes" (Bechtel and Mundale 1999, 176).

35. In the philosophical debate, MR was often taken to imply autonomy in the sense of the denial of intertheoretic reduction (e.g., Fodor 1975, 2000). This sense of autonomy is compatible with as much theorizing using neuroscientific data as a psychologist might want. The sense of autonomy in premise 2, in which psychological theorizing is not constrained at all by developments in neuroscience, is far stronger. Kim (2006, 117) expresses this sense as follows: "Perhaps there might be non-carbon-based or non-protein-based biological organisms with mentality, and we cannot a priori preclude the possibility that nonbiological electromechanical systems, like the 'intelligent' robots and androids in science fiction, might be capable of having beliefs, desires, and even sensations. All this suggests an interesting feature of mental concepts: They seem to carry no constraint on the actual physical-biological mechanisms that, in a given system, realize or implement them."

https://doi.org/10.1086/652964 Published online by Cambridge University Press

neural pathways from all the others. So if premise 3 is intended to make a claim that goes beyond the fact that cognitive neuroscience exists, empirical support for it would have to come from studies showing the essential use of data from a nonhybrid subdiscipline of neuroscience in psychological theorizing.[36] As a matter of fact, however, cognitive psychologists differ sharply regarding the utility, if any, of neuroscientific (in particular, neuroimaging) data in psychological theorizing, quite independently of their positions vis-à-vis MR. For example, Henson (2005; discussed below) defends the use of fMRI data in experimental psychology; Coltheart (2004) defends the 'ultracognitivist' view in cognitive neuropsychology, which sees no use for data regarding where or how functions are realized; and Shallice (1988, 213–14) grants only limited use of anatomical data in cognitive neuropsychology.[37]

If $ANA^{B\&M}$ shows anything, it is that we need a new concept of autonomy appropriate to an empirical MR debate. Since degeneracy is a claim about the relation between psychological and neural hierarchies, this new concept might highlight the difference it makes to our ability to make inferences from knowledge of one hierarchy to hypotheses about the other if the relation between their parts is largely degenerate rather than largely one to one. Obviously, if degeneracy is generally true of a kind of cognitive system, its cognitive and neural hierarchies will not mirror each other. That is, degeneracy is compatible with incommensurate (nonisomorphic) decompositions of functions and structures, while one-to-one mappings imply commensurate (isomorphic) decompositions. So if degeneracy is largely true of a system, we cannot reliably infer from a task analysis to a decomposition of the neural structure that realizes it or vice versa. Conversely, if one-to-one mappings are largely true, these inferences should be reliable. This is a major difference: one-to-one mappings imply that evolved cognitive systems possess a degree of efficiency in realization that would be surprising in evolved structures if not engineered ones. This implication of one-to-one mappings may be the em-

36. By "essential use," I mean that the neuroscientific data must be useful because they are neuroscientific; obviously, scientists can use information from anywhere in their theorizing without implying anything about the relation between the source of that information and their theory: Kekule's dream of a snake eating its tail inspired his model of benzene's ring structure, but that does not make psychological or herpetological information (qua psychological or herpetological) useful in chemistry.

37. See also Ward (2006, 79) on the historical "schism" within cognitive neuropsychology. In computational neuroscience, which develops highly abstract implementations of functional models, Schwartz (1990, x–xii) summarizes a history of disagreement between theoreticians and experimentalists over the relevance of anatomical or other experimental (e.g., behavioral) data to neural net models.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

pirical equivalent of Putnam's (1975, 436) oft-cited claim that the identity theory "is certainly an ambitious hypothesis."

I turn now to claim 2 and Bechtel and Mundale's (1999) methodology-based case against MR. This claim, which is independent of claim 1 and $ANA^{B\&M}$, is inconsistent with the fact that cognitive neuroscientists have proposed the degeneracy hypothesis. It is not entirely clear how Bechtel and Mundale intend claim 2 to be used to argue against MR, but a reasonable reading would embed it in the following inductive "success" argument, which might be called an Argument from Empirical Fruitfulness (AEF):

1. MR claims that the same psychological state can be realized by different brain states.
2. Cognitive researchers assume (implicitly) that MR is actually false (claim 2).
3. Their assuming its falsity has allowed the fruitful use of neuroscientific data in guiding the decomposition and understanding of cognitive systems (Bechtel and Mundale 1999, 204).
4. So (probably) MR is false (or implausible).

Whether Bechtel and Mundale endorse AEF or not, I will only discuss their empirical support for premise 2, without which premise 3 does not get off the ground. A related success argument by Henson, discussed below, will also help show how AEF goes wrong.

Bechtel and Mundale claim (1999, 177) that the denial of MR is implicit in some of the assumptions built into localizationists' research methodologies. These methodologies are such that "(1) the appeal to function, especially psychological function, is an *essential* part of both the [mapping] project and its tools, and (2) the cartographic project itself is frequently carried out comparatively—across species" (177–78).[38] According to Bechtel and Mundale, the implicitly MR-denying assumptions behind 1 include

38. Bechtel and Mundale continue (1999, 177–78): "For multiple realization to be a serious option, brain taxonomy would have to be carried out both independently of psychological function, and without comparative evaluation across species." The first conjunct is ambiguous. If "brain taxonomy" refers to neuroanatomy, then it is carried out independently of cognitive function (even if it is motivated by the desire to explain cognition). If it refers to cognitive neuroanatomy, then it is not carried out independently of cognitive function, and degeneracy cannot imply that it should not be. The second conjunct is discussed in the text. I should note that while Bechtel and Mundale clearly distinguish anatomical and cognitive areas (e.g., "areas delineated by gyri and sulci" and "functionally significant areas"), as well as neurobiology and cognitive neuroscience, they do not disambiguate their term "brain area." Thus, their claim that "brain areas" are individuated partly by the cognitive function they subserve may be expressed unambiguously by saying that the individuation of cognitive areas involves functional and anatomical criteria.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

pure insertion in the experimental paradigm in imaging studies called cognitive subtraction, while the implicitly MR-denying assumption behind 2 is that of *homology*. I will argue that these assumptions do not implicitly deny MR and that their roles in specific experimental paradigms and in cognitive neuroscience generally are compatible with the degeneracy hypothesis. I will start with homology since that concept is more familiar.

According to Bechtel and Mundale (1999), the use of comparative data in cognitive neuroscience relies on assuming cross-species anatomical commonalities: "One might think, at first glance, that the ability to make comparisons across species actually depends upon multiple realizability. In fact, it is the very similarity (or more precisely, homology) of brain structures which permits us to generalize across certain species. So in this latter respect, in the context of neuroscientific research, they are not multiply realized" (177–78). "Most neuroimaging to date is performed on humans while the most detailed neuroanatomical and neurophysiological work (using, e.g., single-cell recording) has been done on other species. As a result, researchers often have to try to coordinate the imaging work on humans with neuroanatomy from other primate species (especially the macaque), and thus are assuming that cognitive functions are not differently realized in the two species" (190). The assumption of anatomical commonalities across species, they argue, corresponds to the way psychologists and philosophers ignore vast cognitive and behavioral differences when they classify (e.g.) fear, pain, or hunger as a single type of psychological state across species: neuroscientists ignore vast physical differences between species in order to classify anatomical areas as being of the same cross-species type.

Bechtel and Mundale are undoubtedly correct about the importance of comparative data in cognitive neuroscience. Brodmann's maps of human and other species' brains were developed from preparations from dozens of mammalian species. Contemporary researchers use cross-species anatomical data as converging evidence for structure-function mappings (Felleman and Van Essen 1987; Kanwisher, McDermott, and Chun 1997; LeDoux 2000) and for human structural imaging (Hagmann et al. 2008). Cognitive-functional imaging data from awake behaving animals (primarily macaques) are increasingly being used as further converging evidence for structure-function mappings (Vanduffel et al. 2001; Orban et al. 2004).

But using comparative data on the basis of presumed homology is not equivalent to typing anatomical areas as the same across species. First, homologies—in this case, brain structures in distinct species that derive from a structure possessed by a common ancestor—are hypothesized, not assumed. What is assumed is the theory of evolution. Evolution justifies and motivates the search for structural, neurophysiological, and cognitive-

https://doi.org/10.1086/652964 Published online by Cambridge University Press

functional homologies between different species.[39] But, second, evolutionary theory does not ipso facto commit researchers to the existence, extent, or nature of specific homologies or to the role that this historical relation may play in the individuation of structural, neurophysiological, or cognitive types or (consequently) structure-function mappings.[40] In particular, it does not commit them to ignoring vast physical differences between possible or actual homologues to classify them as a single cross-species anatomical-area type.[41] Tootell, Tsao, and Vanduffel (2003), who study the macaque visual system, express a concern that justifies this lack of commitment: "The level of accepted evolutionary similarity between possibly homologous cortical regions is likely to decrease (not increase) as we learn more. It is easy to assume that macaque area X is equivalent to human region Y, if almost nothing is known about region Y. However, further study will (by definition) reveal more detailed features, any of which may differ across species" (3986). It is an open question whether in any particular case the similarities will outweigh the known or expected differences when individuating even homologous anatomical areas.[42]

39. Kim (2002) argues that only homoplasies—similar structures that developed independently in species with distinct evolutionary lineages, such as bird and bat wings—count as relevant evidence for or against multiple realizability; Couch (2004) counters that both homologies and homoplasies are relevant. Bechtel and Mundale (1999) refer specifically to homologies, so I set homoplasies aside for the sake of argument.

40. Orban et al. (2004, 317) note: "Cortical areas in humans and macaques are considered homologous if they derive from areas present in a common primate ancestor. For areas that existed in this common ancestor, the challenge is to identify the homologous areas in monkeys and humans despite whatever divergences have occurred in structure, function and geographic location." Functional homologies also are not assumed. For example, it is a matter of debate whether human language derived from a homologous capacity in a primate ancestor or arose de novo via the redeployment of other capacities (Deacon 2004).

41. Trivially, any anatomical area of (e.g.) macaque and human brains can already be classified under the cross-species types "is possessed by a primate," "is derived from a common ancestor," or even "is smaller than a bread box," but presumably these are not the cross-species types localizationists qua localizationists are interested in. Their interest is in properties that ground cognition (or, as Polger [2004, 249 n. 13] puts it, those which "ground the regularities of psychological explanation"). It follows that not every cross-species property, physical or otherwise, counts as a cross-species realizer, on pain of trivializing the debate. This suggests that Bechtel and Mundale's diagnosis of the intuitive appeal of MR is inadequate: even if philosophers did mismatch taxonomic levels, thereby making MR appear more plausible than it is, Bechtel and Mundale err in thinking that any one-to-one mapping affects the MR debate equally.

42. Thus, many if not all references to (e.g.) "primate visual cortex" are most plausibly construed as references to species-specific cognitive areas that (i) are possessed by primates and (ii) as a working hypothesis involve functions that are similar across species.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

Moreover, even granting for the sake of argument that a well-confirmed homology between anatomical areas commits neuroscientists to classifying these areas under the same cross-species type, it is a separate step to the conclusion that the cognitive functions assigned to these anatomical areas will be the same (let alone homologous). Vision research, which provides some of the best empirical support we have for cross-species anatomical-area types so far, shows that some structure-function mappings are identical across species and others are not. In the macaque, visual area V3 is moderately motion and direction sensitive, while V3A is not; in the human, the functions performed by these areas are reversed (Felleman and Van Essen 1987; Tootell et al. 1997, 2003; Vanduffel et al. 2001; Orban et al. 2004; Vanduffel et al. [2002] discuss other macaque-human disanalogies). So even anatomical areas that are homologous and (by assumption) cross-species-type identical can be mapped to distinct functions.[43]

These prima facie cases of MR are based on the same methodologies that result in prima facie one-to-one mappings across species (e.g., V1 in macaques and humans may be an example). Nor should the V3/V3A case be considered exceptional. Tootell et al. (2003) also raise a general concern for future mappings in visual areas: "Thus, the retinotopy defining a region is not absolutely linked to the functional properties of the same region. When such properties differ, do we assume homology based on the retinotopic criteria or the functional criteria? Here the retinotopy appears more fundamental (conserved)" (3982). That is, in the V3/V3A case, retinotopy determines which cognitive areas (e.g., V3) are typed as the same across species, and the function of motion sensitivity is mapped to distinct cognitive areas in each species.[44] Thus, shared anatomical (or cognitive) areas and functions do not entail shared structure-function mappings. In general, where multiple criteria are used in individuation, in cases of conflict the taxonomic result cannot be determined a priori. And there

43. Tootell et al. (1997, 7075) remark: "This distinction [between V3 and V3A in humans and macaques regarding motion sensitivity] is so salient that one anonymous reviewer suggested switching the (retinotopically based) names V3 and V3A in humans to match the 'reversal' of motion selectivity in humans. Although this suggestion has merit, we have chosen to leave the human names V3 and V3A consistent with their retinotopic counterparts in the macaque, implicitly assuming that the motion selectivity has become wired differently in V3 and V3A in monkeys, as compared to humans."

44. This introduces additional complexity to the notion of a structure-function mapping: cognitive areas can be individuated (in part) by reference to one cognitive criterion (organization of visual field) and have other (nonindividuating) cognitive functions assigned to them.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

are plenty of cases with potential for conflict.[45] After all, simpler functions are generally easier to realize in more than one kind of structure (Kary and Mahner 2002). Cognitive functions simple enough to be typed across species (e.g., motion detection) may be the psychological equivalent of McJobs, in the same way that many different kinds of things can uncork a wine bottle or be an "and" gate (or man a fry station).

But Bechtel and Mundale (1999) also claim to find support for claim 2 from the methodologies used in lesion (or deficit) studies and imaging studies (typically involving human subjects): "In interpreting [lesion-based cognitive] deficits, researchers implicitly reject multiple realization among human brains and assume that damage to a brain area in anyone will result in a deficit to a particular cognitive function that is performed by that area in undamaged brains" (184). "Many [imaging] researchers employ the *subtractive method* to focus on brain activities associated with component psychological processes." After describing this method, which I explain below, they continue: "Identifying brain areas through neuroimaging depends critically on the cognitive tasks subjects are asked to perform; thus, the possibility of multiple realizability is restricted at the outset" (190). The latter quotation refers to the pure insertion assumption used in cognitive subtraction. The former quotation is consistent with three different assumptions—transparency (in lesion studies), universality, and unipotentiality—that Bechtel and Mundale do not explicitly introduce. I will describe these but focus on pure insertion.

Bechtel and Mundale also note that it is standard in imaging studies to transform individual-subject results, voxel by voxel, into the coordinates of a standard brain map and to average results across subjects. That these procedures leave common areas of statistically significant activation across subjects, Bechtel and Mundale (1999, 190) argue, "suggests much less variability than the multiple realizability arguments would allow." However, these procedures are as worrisome to many researchers as methodological assumptions because of significant individual differences that are lost on averaging (see fig. 3).

Before turning directly to Bechtel and Mundale's claims, it is instructive

45. For example, although humans and macaques both have ventral (object recognition, "what") and dorsal (spatial location, "where") pathways after V1, the object-recognition stream in humans, unlike in the macaque, runs almost entirely on the ventral surface of the temporal lobe and does not extend as far forward, while the spatial-location stream runs along a more superior route in parietal cortex in humans than in the macaque (Haxby et al. 1991; Ungerleider, Courtney, and Haxby 1998). Ungerleider et al. (1998) also propose a dual-pathway model for working memory, with object working memory in ventrolateral prefrontal cortex and spatial working memory in dorsolateral prefrontal cortex. But their postulated human area for spatial working memory is located both more superior and more posterior to that of the macaque.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

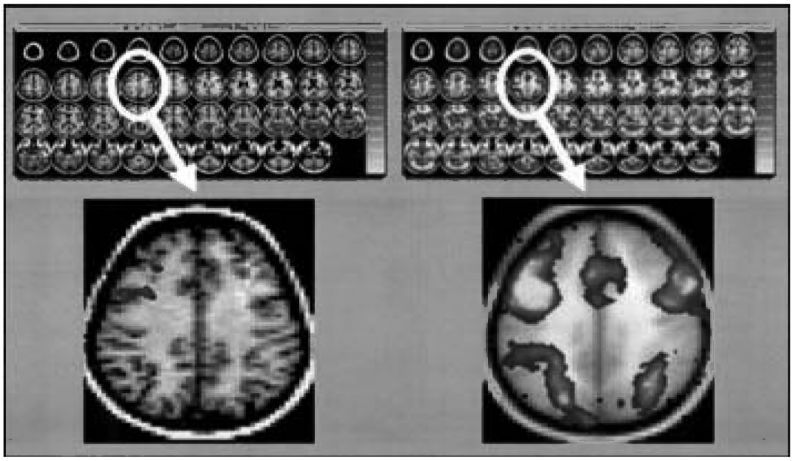


Figure 3. fMRI study of normals engaged in a semantic-processing task. Enlarged images are of a Talaraich *z*-level slice in one subject (*left*) and averaged across 12 subjects (*right*). Source: Savoy (2001, 29). Color version available as an online enhancement.

to compare AEF with a somewhat similar "success" argument from Henson (2005). Henson argues that a "working hypothesis" of "strong systematicity" ("a one-to-one mapping between functional and structural units"; 193) is necessary in experimental psychology in order to justify the use of imaging data to inform psychological models—specifically, to justify inductions from the activation of a structure associated with a function in one experimental paradigm to the performance of the same function in other experimental contexts in which that structure is activated. But to assume strong systematicity is not to assume MR is false; Henson explicitly acknowledges the possibility of degeneracy within and across subjects. The assumption, rather, is that an anatomical region (or network) is not pluripotential: "For inferences of the 'structure-function induction' type . . . one must assume a one-to-one mapping between function and structure, in order to discount the possibility that the same structure implements different functions across experimental contexts in which the [structure-to-function] inference is made" (220).[46] As noted, unipotentiality is compatible with degeneracy, as it does not rule out latent

46. Kosslyn (1999, 1290) raises the same concern about the validity of structure-to-function inferences given the possibility of pluripotentiality: even for sensory and motor cortices, "we simply do not know enough to exclude the possibility of multiple roles for any piece of cortical real estate."

https://doi.org/10.1086/652964 Published online by Cambridge University Press

degenerate regions (networks) or overlapping degenerate networks within subjects or degeneracy across subjects. Moreover, Henson adds that the assumption of strong systematicity "cannot be proved on independent grounds and is probably best evaluated by the success of the [localization] enterprise as a whole." This is because if we do not already know (or hypothesize) the functional organization of the mind when doing imaging studies, we cannot know (hypothesize) that an activated area has been associated with just the right function, at the appropriate level of generalization. We must assume we have got it right and "see how much progress is made" (219). In short, Henson argues, if we assume unipotentiality (not the falsity of MR), imaging data can be of use in developing cognitive models (not that this use is already fruitful), with the hope that partly as a result of using these data a consistent functional hierarchy will emerge. If so, then the assumption would be justified.

Bechtel and Mundale's argument for claim 2, however, rests largely on the methodology of cognitive subtraction used in imaging studies. Since the whole brain is always active, localization of function using imaging data requires measuring relative, not absolute, differences in neural activity. The method of cognitive subtraction involves comparing activity measured during a baseline (or comparison) task and activity in a closely matched experimental (or activation) task that is thought to differ from the baseline task by one component.[47] For example, the baseline task may be passively seeing a written word, and the experimental task is seeing a written word and saying it aloud. Activity imaged during the baseline task is "subtracted" from that imaged during the experimental task. (Often only the peaks or centers of mass of statistically significant areas of activation left after subtraction—and after averaging and smoothing procedures—are shown in published imaging results.) The function has been localized when we infer that the new cognitive component in the experimental task is subserved by the region(s) of additional activation remaining after subtraction.

The legitimacy of this inference depends on the assumption of pure

47. Baseline activity is neural activity above a chosen threshold of statistical significance detected during performance of the baseline task; activity detected in the experimental condition is relative to the same threshold. Developing a baseline that involves all and only the same processes as the experimental task (except the process of interest) is problematic (Price and Friston 1997), as is determining an appropriate threshold for statistical significance, since depending on the choice of threshold, differing areas of activity may be implicated in subserving a cognitive function. To borrow Savoy's (2001, 28–30) analogy: if your task is to count islands, how many you find and how large they are will depend on the sea level.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

insertion.[48] Whenever a task is added, there is processing involved due to the interaction between the new and the old tasks. The pure insertion assumption is that the amount of additional activity attributable to the interaction is zero. At the psychological level, the assumption is that cognition proceeds via isolable steps (in the simplest cases, in a serial, feed-forward manner), such that adding a task has no effect on the way previous tasks are performed. At the neural level, the assumption is that the difference in neural activity imaged during the baseline and experimental tasks is due entirely to the new task and does not represent any influences on or interaction with the baseline activity. In short, the pure insertion assumption is that both functional and structural systems are modular.[49] Pure insertion rules out the possibility that the additional neural activity may be neither sufficient nor necessary for the presumed-to-be purely additional task, thereby making an inference from a one-to-one (new task : new activity) mapping possible.

An analogous assumption in cognitive neuropsychology is called the transparency assumption (Caramazza 1986). Cognitive neuropsychologists study patterns of cognitive deficits to develop models of normal cognitive function that can be mapped to the lesioned areas. (Double dissociations are patterns of deficit considered especially helpful; see n. 21.) However, the legitimacy of using lesioned subjects (e.g., stroke victims) to understand normal cognition must be justified. The transparency assumption claims that the same cognitive system exists through change; there is not a prelesion system and a distinct reorganized postlesion system. Transparency rules out the possibility that the lesion site subserves part of a new reorganized system rather than the specific deficit(s). This assumption has also been questioned (see Ward 2006, 85) but is best examined in the context of a discussion of plasticity, which I set aside here.[50]

When researchers average the imaging results from multiple subjects and plot them to a common brain atlas to identify areas of statistically

48. The assumption of pure insertion originates in experimental psychology (Sternberg 1969), where it is part of the "additive factors" method of testing hypotheses about processing stages on the basis of differences in reaction times when comparing baseline and experimental tasks. This assumption is still widely used in experimental psychology (Henson 2005, 212) but not in imaging studies (Ward 2006, 56–62).

49. This is modularity without the usual bells and whistles (Fodor 1983; Carruthers 2006). I cannot go into the modularity debate here, but a skeptic about modularity should be skeptical about pure insertion.

50. "Plasticity" refers to the brain's ability to change in response to experience. This includes both its development over a lifetime (but especially in childhood) and the restitution of function after damage. Plasticity has been used to argue for MR in existing biological organisms (see Shapiro [2004] for a detailed response), but for space reasons this debate cannot be adequately assessed here.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

significant activity across subjects, they also assume universality: that all (human) cognitive systems are basically the same (Caramazza 1986). This assumption legitimizes generalizing from a small experimental sample to a larger group. Since it equally legitimizes inferring from degeneracy in a sample to degeneracy in a larger group, it cannot tell against degeneracy.

Cognitive subtraction, and the assumption of pure insertion on which it rests, has enabled researchers to use imaging data to begin to unravel the mystery of brain-based cognition. But pure insertion is undermined by research indicating nonnegligible interactions between areas, such as in cases of effective connectivity and the influence of directed attention (Friston 1997), the effect of practice (Raichle et al. 1994), and the relation between the current task(s) and what the subject has just done before testing (Kosslyn 1999, 1291–92; see also Pachella 1974 and Uttal 1998). In these cases, the regions associated with the tasks shared in the baseline and experimental conditions differ when a new processing component is added. For example, in a direct test of pure insertion, Jennings et al. (1997) found that the areas associated with the baseline task differed depending on how subjects were asked to respond. Subjects were given a semantic task (determining whether a word represents something living) and a letter-judgment task (determining whether a word contained the letter "a") and, in both cases, were asked to give a yes or no response in three ways: silent thought, clicking a mouse, or answering aloud. The pair-wise subtractions were between (e.g.) reading, making a semantic judgment, and responding yes or no by clicking a mouse versus reading, making a semantic judgment, and responding yes or no by saying yes or no aloud. If pure insertion were true, the areas associated with the semantic task should remain the same across response modes. Instead, the active areas differed significantly. Although the left inferior frontal cortex was activated in all semantic-processing conditions (albeit to different degrees, depending on the response condition), other areas were unique to semantic processing given a particular response mode or to only two of them. And even the existence of a common area of activation is consistent with the hypothesis that this region is a necessary part of multiple degenerate networks for semantic processing.

In imaging studies, experimental paradigms that avoid pure insertion are now favored. For example, in the cognitive conjunction paradigm (Price and Friston 1997), which is a modification of cognitive subtraction, several task pairs are developed that share only the cognitive component of interest.[51] Since activity due to interaction will be specific to each pair,

51. Other methods include using factorial and parametric designs (Friston 1997). In a factorial design, the aim is to identify areas of interaction explicitly by comparing activation during performance of tasks combined as factors or variables (e.g., if P =

https://doi.org/10.1086/652964 Published online by Cambridge University Press

any activity that remains across subtractions can be associated with the component of interest. In their PET study, they imaged subjects performing subtraction tasks that shared only the process of phonological retrieval from visually presented stimuli (defined as activating the name attached to a visual stimulus or concept). For example, one pair-wise subtraction was reading a visually presented word and saying a prespecified word to visually presented strings of false font; another was naming a visually presented Arabic letter and saying the same prespecified word to single false-font characters. Significant activation surviving subtraction across all task pairs was found in the left posterior basal temporal lobe, left frontal operculum, and midline cerebellum. It does not follow that phonological retrieval can be localized to these regions. Because of the possibility of degeneracy, Price and Friston (2002) themselves deny the legitimacy of inferring from the necessity of an activated area (or areas) for the component of interest. The common area(s) may not be necessary (e.g., if there is latency) or sufficient (e.g., if it participates in multiple degenerate networks). Nevertheless the data are valuable for helping to determine the role(s) of the activated cortical areas without presuming which structure-function theory may eventually prove correct.

In addition, Noppeney et al. (2004) have proposed a new methodology designed specifically to find evidence of degeneracy within subjects (see fig. 4). In their iterated model, normal subjects are imaged to identify candidate regions of interest and generate hypotheses about potential degenerate neural systems. Lesions to these regions (in patients or temporarily induced in normals) then test whether a specific behavioral deficit occurs after damage to one region (prima facie evidence of its necessity) or more than one (prima facie evidence of degeneracy). If lesioning does not yield a deficit, a latent degenerate system may be operating. Imaging patients with lesions in the original regions of interest but who can perform the task at normal levels can reveal these latent systems.

To summarize, cognitive neuroscience methodologies do not rely even implicitly on the assumption that MR is false. Nor do the revised and new methods presuppose that degeneracy is true. They aim to provide better data from which to infer more reliably whichever of the competing theories is more explanatory.

It is also worth noting that the validity of inferences made in imaging studies in particular are likely to be affected by technological developments. One, already noted, is the increasing use of diffusion imaging to

phonological retrieval and O = object recognition, a factorial design would evaluate activity during conditions P&O, ~P&O, P&~O, and ~P&~O). Parametric designs treat variables as dimensions (i.e., with degrees of activation) rather than categories (i.e., active or not). See also Ward (2006, 56–62).

https://doi.org/10.1086/652964 Published online by Cambridge University Press

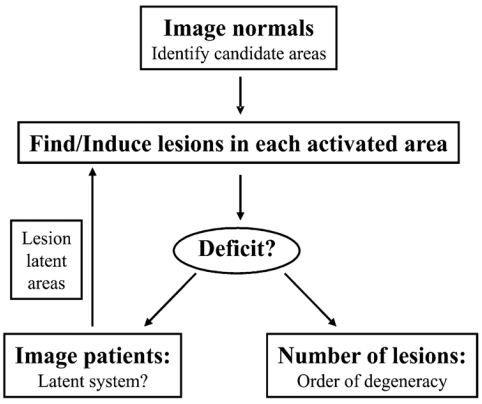


Figure 4. Proposed method for identifying degenerate neuronal systems by combining functional imaging and neuropsychological lesion data from normal subjects and lesion patients in an iterative procedure. Bottom left arrow (after "Deficit?") is no, and the bottom right arrow is yes. Source: Noppeney, Friston, and Price (2004, 438).

reveal axonal connectivity, not just activated regions, in normal human subjects. Another is the recently authorized use on human subjects of fMRI machines that can generate magnetic fields of up to 8 tesla. Most fMRI machines in current experimental use operate at 1 or 1.5 tesla, a level at which just one in 100,000 hydrogen atoms in the subject's brain aligns with the generated magnetic field.[52] The stronger machines raise an interesting problem (Savoy 2001, 30–31). Even when keeping the choice of threshold, experimental paradigm, and method of data analysis fixed, additional statistically significant activation is likely to be detected with stronger fMRI machines. It follows that at least some regions of activation (and current cortical cognitive areas) may be artifacts of the power of most fMRI machines used experimentally up to now. Not only are identified areas likely to morph and grow; some are likely to be statistically significantly active (if not maximally so) during other tasks at 8 tesla, even if this activity is below threshold at 1.5 tesla. In other words, an

52. In fMRI, the aligned hydrogen nuclei are temporarily knocked 90 degrees off alignment by an electrical pulse; their movement back to alignment generates the signal from which we infer statistically significant differences in blood oxygenation levels between brain regions and from there to significant differences in neural activity, which may in turn be associated with a task that the imaged subject performs.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

area that "looks" unipotential at 1.5 tesla may turn out to be clearly pluripotential at 8 tesla.[53]

5. Conclusion. I have argued that (i) MR is a live empirical hypothesis within cognitive neuroscience and (ii) some of the main cognitive neuroscientifically based arguments against the plausibility of MR fail. There is no basis for claiming (as Bechtel and Mundale [1999] do) that the empirical plausibility of MR "evaporates" in the light of neuroscientific research. To the contrary, it is a genuine mystery why so much of the philosophical literature on the implications of cognitive neuroscience for MR has been so negative in its conclusions. As far as I can tell, Putnam's intuitions may ultimately be vindicated for evolved biological organisms, without even considering what may be the case for Martians and robots. At the very least, supporters of MR, and of nonreductive physicalism, have no reason to shy away from defending their positions on cognitive neuroscientific grounds, leaving intuition and analogy far behind.

REFERENCES

Aizawa, Kenneth, and Carl Gillett. 2009. "Levels, Individual Variation, and Massive Multiple Realization in Neurobiology." In *The Oxford Handbook of Philosophy and Neuroscience*, ed. John Bickle, 539–81. New York: Oxford University Press.

Andrews, Kristin. 2008. *Stanford Encyclopedia of Philosophy*, s.v. "Animal Cognition," http://plato.stanford.edu/entries/cognition-animal/.

Attwell, David, and Costantino Iadecola. 2002. "The Neural Basis of Functional Brain Imaging Signals." *Trends in Neurosciences* 25 (12): 621–25.

Bechtel, William. 2006. "Critical Notice: The Mind Incarnate." *Philosophy and Phenomenological Research* 73 (2): 497–500.

Bechtel, William, and Jennifer Mundale. 1999. "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* 66:175–207.

Bickle, John. 1998. *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.

———. 2003.*Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer.

Block, Ned J., and Jerry A. Fodor. 1972. "What Psychological States Are Not." *Philosophical Review* 8 (2): 159–81.

Broca, Paul. 1861/1960/2001. "Remarques Sur le Siege de la Faculte du Langage Articule; Survies d'une Observation d'Aphemie." *Bulletin de la Societe Anatomique* 6:330–57. Trans. G. von Bonin. 1960. "Remarks on the Seat of the Faculty of Articulate Language, Followed by an Observation of Aphemia." In *Some Papers on the Cerebral Cortex*, ed. G. von Bonin, 49–72. Springfield, IL: Thomas. Repr. in *Philosophy and the Neurosciences: A Reader*, ed. William Bechtel, Pete Mandik, Jennifer Mundale, and Robert Stufflebeam, 87–99. Malden: Blackwell.

Brodmann, Korbinian. 1909/1994. *Vergleichende Lokalisationslehre der Grosshirnrinde*. Leipzig: Barth. Trans. L. J. Garey. 1994. *Brodmann's Localisation in the Cerebral Cortex*. London: Smith-Gordon.

53. Also, and in contrast with Bechtel and Mundale's claim that averaged imaging data provide evidence of commonality, Savoy (2001, 31) considers averaged data a harbinger of this problem. In images showing averaged data from various subjects, the areas of maximal activity tend to be both more numerous and larger.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

Burton, H. 2003. "Visual Cortex Activity in Early and Late Blind People." *Journal of Neuroscience* 23 (10): 4005–11.
Caramazza, Alfonso. 1986. "On Drawing Inferences about the Structure of Normal Cognitive Systems from the Analysis of Patterns of Impaired Performance: The Case for Single-Patient Studies." *Brain and Cognition* 5:41–66.
Carruthers, Peter. 2006. *The Architecture of the Mind*. New York: Oxford University Press.
Churchland, Patricia. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.
Churchland, Paul. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78 (2): 67–90.
Clapp, Leonard. 2001. "Disjunctive Properties: Multiple Realizations." *Journal of Philosophy* 98 (3): 111–36.
Coltheart, Max. 2004. "Brain Imaging, Connectionism, and Cognitive Neuropsychology." *Cognitive Neuropsychology* 21 (1): 21–25.
Coltheart, Max, and Martin Davies. 2003. "Inference and Explanation in Cognitive Neuropsychology." *Cortex* 39:188–91.
Couch, Mark B. 2004. "Discussion: A Defense of Bechtel and Mundale." *Philosophy of Science* 71:198–204.
Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.
Deacon, Terrence. 2004. "Monkey Homologues of Human Language Areas: Computing the Ambiguities." *Trends in Cognitive Sciences* 8 (7): 288–90.
Dronkers, Nina F., Brenda B. Redfern, and Robert T. Knight. 2000. "The Neural Architecture of Language Disorders." In *The New Cognitive Neurosciences*, ed. Michael Gazzaniga, 949–58. Cambridge, MA: MIT Press.
Dunn, John C., and Kim Kirsner. 2003. "What Can We Infer from Double Dissociations?" *Cortex* 39:1–7.
Edelman, Gerald M., and Joseph A. Gally. 2001. "Degeneracy and Complexity in Biological Systems." *Proceedings of the National Academy of Sciences of the USA* 98 (24): 13763–68.
Farah, Martha J. 1994. "Neuropsychological Inference with an Interactive Brain: A Critique of the 'Locality' Assumption." *Behavioral and Brain Sciences* 17 (1): 43–104.
Felleman, Daniel J., and David C. Van Essen. 1987. "Distributed Hierarchical Processing in the Primate Cerebral Cortex." *Cerebral Cortex* 1:1–47.
———. 1991a. "Distributed Hierarchical Processing in the Primate Cerebral Cortex." *Cerebral Cortex* 1:1–47.
———. 1991b. "Receptive Field Properties of Neurons in Area V3 of Macaque Monkey Extrastriate Cortex." *Journal of Neurophysiology* 57:889–920.
Field, Hartry. 1973. "Theory Change and the Indeterminacy of Reference." *Journal of Philosophy* 70:462–81.
Fodor, Jerry. 1974/1975. "Special Sciences; or, The Disunity of Science as a Working Hypothesis." *Synthese* 28:97–115. Rev. and repr. in *The Language of Thought*, 9–26. Cambridge, MA: Harvard University Press.
———. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
———. 2000. "Special Sciences: Still Autonomous after All These Years; A Reply to Jaegwon Kim's 'Multiple Realization and the Metaphysics of Reduction.'" In *Critical Condition*, 9–24. Cambridge, MA: MIT Press. Repr. from *Philosophical Perspectives 11: Mind, Causation, and World*, ed. J. Tomberlin. Atascadero, CA: Ridgeview.
Friston, Karl. 1997. "Imaging Cognitive Anatomy." *Trends in Cognitive Sciences* 1 (1): 21–27.
———. 2002. "Beyond Phrenology: What Can Neuroimaging Tell Us about Distributed Circuitry?" *Annual Review of Neuroscience* 25:221–50.
Friston, Karl, and Cathy Price. 2003. "Degeneracy and Redundancy in Cognitive Anatomy." *Trends in Cognitive Sciences* 7 (4): 151–52.
Gillett, Carl. 2003. "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences." *Journal of Philosophy* 100 (11): 591–603.
Gregory, R. L. 1961. "The Brain as an Engineering Problem." In *Current Problems in Animal*

https://doi.org/10.1086/652964 Published online by Cambridge University Press

Behaviour, ed. W. H. Thorpe and O. L. Zangwill, 307–30. London: Cambridge University Press.
Grodzinsky, Yosef, and Andrea Santi. 2008. "The Battle for Broca's Region." *Trends in Cognitive Sciences* 12 (12): 474–80.
Hagmann, Patric, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J. Honey, Van J. Wedeen, and Olaf Sporns. 2008. "Mapping the Structural Core of Human Cerebral Cortex." *Public Library of Science (PLoS) Biology* 6 (7): 1–15.
Haxby, James V., Cheryl I. Grady, Barry Horwitz, Leslie G. Underleider, Mortimer Mishkin, Richard E. Carson, Peter Herscovitch, Mark B. Schapiro, and Stanley I. Rapoport. 1991. "Dissociation of Object and Spatial Visual Processing Pathways in Human Extrastriate Cortex." *Proceedings of the National Academy of Sciences of the USA* 88 (5): 1621–25.
Henson, Richard. 2005. "What Can Functional Neuroimaging Tell the Experimental Psychologist?" *Quarterly Journal of Experimental Psychology* 58A (2): 193–233.
Jennings, Janine M., Anthony R. McIntosh, Shitij Kapur, Endel Tulving, and Sylvain Houle. 1997. "Cognitive Subtractions May Not Add Up: The Interaction between Semantic Processing and Response Mode." *Neuroimage* 5:229–39.
Johnson, Adam, Andre A. Fenton, Cliff Kentros, and A. David Redish. 2009. "Looking for Cognition in the Structure within the Noise." *Trends in Cognitive Sciences* 13 (2): 55–64.
Kanwisher, Nancy, Josh McDermott, and Marvin Chun. 1997. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception." *Journal of Neuroscience* 17 (1): 4302–11.
Kary, Michael, and Martin Mahner. 2002. "How Would You Know if You Synthesized a Thinking Thing?" *Minds and Machines* 12:61–86.
Keeley, Brian L. 2000. "Shocking Lessons from Electric Fish: The Theory and Practice of Multiple Realization." *Philosophy of Science* 67:444–65.
———. 2002. "Making Sense of the Senses: Individuating Modalities in Humans and Other Animals." *Journal of Philosophy* 99 (1): 5–28.
Kim, Jaegwon. 1993. "Multiple Realization and the Metaphysics of Reduction." In *Supervenience and Mind: Selected Philosophical Essays*, 309–35. New York: Cambridge University Press.
———. 2006. *Philosophy of Mind*. 2nd ed. Cambridge, MA: Westview.
Kim, Sungsu. 2002. "Testing Multiple Realizability: A Discussion of Bechtel and Mundale." *Philosophy of Science* 69:606–10.
Kosslyn, Stephen M. 1999. "If Neuroimaging Is the Answer, What Is the Question?" *Philosophical Transactions of the Royal Society* B 354 (1387): 1283–94.
Kosslyn, Stephen M., Nathaniel M. Alpert, William L. Thompson, Christopher F. Chabris, Scott L. Rauch, and Adam K. Anderson. 1994. "Identifying Objects Seen from Different Viewpoints: A PET Investigation." *Brain* 117:1055–71.
Kosslyn, Stephen M., Gregory J. DiGirolamo, William L. Thompson, and Nathaniel M. Alpert. 1998. "Mental Rotation of Objects versus Hands: Neural Mechanisms Revealed by Positron Emission Tomography." *Psychophysiology* 35:151–61.
Lashley, Karl S. 1929. *Brain Mechanisms and Intelligence*. Chicago: University of Chicago Press.
———. 1950. "In Search of the Engram." *Symposia of the Society for Experimental Biology* 4:454–82.
LeDoux, Joseph E. 2000. "Emotion Circuits in the Brain." *Annual Review of Neuroscience* 23:155–84.
Le Poidevin, Robin. 2000. "Space and the Chiral Molecule." In *Of Minds and Molecules: New Philosophical Perspectives on Chemistry*, ed. Nalini Bhushan and Stuart Rosenfeld, 129–41. Oxford: Oxford University Press.
Marshall, John C., and Gereon R. Fink. 2003. "Cerebral Localization, Then and Now." *NeuroImage* 20:S2–S7.
Mundale, Jennifer. 2002. "Concepts of Localization: Balkanization in the Brain." *Brain and Mind* 3:1–18.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

Noppeney, Uta, Karl Friston, and Cathy Price. 2004. "Degenerate Neuronal Systems Sustaining Cognitive Functions." *Journal of Anatomy* 205:433–42.
Orban, Guy A., David Van Essen, and Wim Vanduffel. 2004. "Comparative Mapping of Higher Visual Areas in Monkeys and Humans." *Trends in Cognitive Sciences* 8 (7): 315–24.
Pachella, Robert G. 1974. "The Interpretation of Reaction Time in Information-Processing Research." In *Human Information Processing: Tutorials in Performance and Cognition*, ed. Barry H. Kantowitz, 41–82. Hillsdale, NJ: Erlbaum.
Phillips, C. G., S. Zeki, and H. B. Barlow. 1984. "Localization of Function in the Cerebral Cortex: Past, Present and Future." *Brain* 104 (1): 328–61.
Plaut, David C. 1995. "Double Dissociation without Modularity: Evidence from Connectionist Neuropsychology." *Journal of Clinical and Experimental Neuropsychology* 17 (2): 291–321.
Polger, Thomas. 2004. *Natural Minds*. Cambridge, MA: MIT Press.
Price, Cathy J., and Karl J. Friston. 1997. "Cognitive Conjunction: A New Approach to Brain Activation Experiments." *NeuroImage* 5:261–70.
———. 2002. "Degeneracy and Cognitive Anatomy." *Trends in Cognitive Sciences* 6 (10): 416–21.
———. 2005. "Functional Ontologies for Cognition: The Systematic Definition of Structure and Function." *Cognitive Neuropsychology* 22 (3): 262–75.
Putnam, Hilary. 1967. "The Nature of Mental States." In *Mind, Language and Reality: Philosophical Papers*, vol. 2, 429–40. New York: Cambridge University Press. Repr. "Psychological Predicates." In *Art, Mind and Religion*, ed. W. H. Capitan and Daniel D. Merrill, 37–48. Pittsburgh: University of Pittsburgh Press.
———. 1975. "Philosophy and Our Mental Life." In *Mind, Language and Reality: Philosophical Papers*, vol. 2, 291–303. New York: Cambridge University Press.
Raichle, Marcus E., Julie E. Fiez, Tom O. Videen, Ann-Mary K. MacLeod, Jose V. Pardo, Peter T. Fox, and Steven E. Petersen. 1994. "Practice-Related Changes in Human Brain Functional Anatomy during Nonmotor Learning." *Cerebral Cortex* 4 (1): 8–26.
Savoy, Robert L. 2001. "History and Future Directions of Human Brain Mapping and Functional Neuroimaging." *Acta Psychologica* 107:9–42.
Schwartz, Eric L., ed. 1990. *Computational Neuroscience*. Cambridge, MA: MIT Press.
Shallice, Tim. 1988. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
Shapiro, Lawrence A. 2000. "Multiple Realizations." *Journal of Philosophy* 97 (12): 635–54.
———. 2004. *The Mind Incarnate*. Cambridge, MA: MIT Press.
Sternberg, Saul. 1969. "The Discovery of Processing Stages: Extensions of Donders' Method." *Acta Psychologica* 30:276–315.
Talairach, J., and P. Tournoux. 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical.
Tononi, Giulio, Olaf Sporns, and Gerald M. Edelman. 1999. "Measures of Degeneracy and Redundancy in Biological Networks." *Proceedings of the National Academy of Sciences of the USA* 96:3257–62.
Tootell, Roger B. H., Janine D. Mendola, Nouchine K. Hadjikhani, Partick J. Ledden, Arthur K. Liu, John B. Reppas, Martin I. Sereno, and Anders M. Dale. 1997. "Functional Analysis of V3A and Related Areas in Human Visual Cortex." *Journal of Neuroscience* 17 (18): 7060–78.
Tootell, Roger B. H., Doris Tsao, and Wim Vanduffel. 2003. "Neuroimaging Weighs In: Humans Meet Macaques in 'Primate' Visual Cortex." *Journal of Neuroscience* 23 (10): 3981–89.
Ungerleider, Leslie, Susan Courtney, and James V. Haxby. 1998. "A Neural System for Human Visual Working Memory." *Proceedings of the National Academy of Sciences of the USA* 95:883–90.
Ungerleider, Leslie, and Mortimer Mishkin. 1982. "Two Cortical Visual Systems." In *Analysis of Visual Behavior*, ed. David J. Ingle, Melvyn A. Goodale, and Richard J. W. Mansfield, 549–86. Cambridge, MA: MIT Press.

https://doi.org/10.1086/652964 Published online by Cambridge University Press

Uttal, William R. 1998. *Toward a New Behaviorism: The Case against Perceptual Reductionism*. Mahwah, NJ: Erlbaum.

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

Vanduffel, W., D. Fize, J. B. Mandeville, K. Nelissen, P. Van Hecke, B. R. Rosen, R. B. Tootell, and G. A. Orban. 2001. "Visual Motion Processing Investigated Using Contrast Agent-Enhanced fMRI in Awake Behaving Monkeys." *Neuron* 32:565–77.

Vanduffel, W., D. Fize, H. Peuskens, K. Denys, S. Sunaert, J. T. Todd, and G. A. Orban. 2002. "Extracting 3D from Motion: Differences in Human and Monkey Intraparietal Cortex." *Science*, n.s., 298 (5592): 413–15.

Von Melchner, Laurie, Sarah Pallas, and Mrisanka Sur. 2000. "Visual Behaviour Mediated by Retinal Projections Directed to the Auditory Pathway." *Nature* 404:871–76.

Ward, Jamie. 2006. *The Student's Guide to Cognitive Neuroscience*. Hove: Psychology.

https://doi.org/10.1086/652964 Published online by Cambridge University Press