

About Face:
Philosophical Naturalism, the Heuristic Identity Theory,
and Recent Findings about Prosopagnosia*

Robert N. McCauley
Center for Mind, Brain, and Culture
Emory University

1. *Introduction*

It was a scientist, Ullin Place (1956), who first proposed the modern psycho-physical identity theory. Over subsequent decades, though, it has been philosophers of mind, who have seized on Place's proposal. Those philosophers directed most of their energies at finding fault with the identity theory. Armed with logic, ordinary language, an endless supply of intuitions about mind and mentality, thought experiments, and attention to the accomplishments of computers, philosophers -- functionalists especially -- assembled a list of arguments against the identity theory, most of which, however, fall into one or the other of two categories.

The first group involves various apparent failures of the putative identities to satisfy Leibniz's law with regard to such matters as spatial properties, representational contents, consciousness, etc. Arguments about the identity theory's putative explanatory gaps and about its inability to surpass correlations are the most popular formulations of the first group's complaints. Broadly, the arguments hold either that the identity theory fails to explain how the brain can be said, sensibly, to possess various psychological properties or that any (conceivable) evidence for an identity of mind and brain can establish no more than correlations between them.

By contrast, arguments of the second sort do not fret about the identity theory's alleged explanatory failures. Instead, the general strategy is to burden the identity theory with an embarrassment of riches. Multiple realizability arguments point to the prospect of so many possible physical arrangements realizing psychological states as to make the identity theory appear parochial in its focus on the *human* brain.

William Bechtel and I have argued in two joint papers (Bechtel and McCauley 1999 and McCauley and Bechtel 2001) and independently (McCauley 2007 and Bechtel 2008) for a version of the identity theory, which we have dubbed Heuristic Identity Theory (HIT), that avoids both of these sorts of objections and squares better with actual scientific practice than earlier versions of the identity theory. We both see HIT as but one component of a larger naturalist program in philosophy that maintains that, all things being equal, philosophical projects in general are pursued more responsibly when they are pursued in the light of the activities, the methods, and the findings of the empirical sciences that philosophy has spawned both in centuries past and, in the cases of the psychological and cognitive sciences, fairly recently. (See, for example, Thagard 2010.) Exploring the identity theory within the framework of the sciences that are the most relevant to considering the relations of psychological and neural phenomena, rather than exploring it in splendid philosophical isolation, yields a far more sanguine view of its prospects.

Section 2 lays out the naturalist's case for assessing the merits of any version of the identity theory primarily with respect to how well it harmonizes with the activities and findings

of the relevant sciences. Section 3 outlines HIT, exhibits its implicit commitment to explanatory pluralism, and shows how HIT manages the two sorts of objections to the identity theory noted above. Section 4 sketches how a current program of scientific research illustrates HIT in action. It explores the interplay between hypothetical identities and empirical findings in recent psychological science and cognitive neuroscience concerning the neural realization of human beings' abilities to detect and identify human faces.

2. *A Naturalistic Take on Assessing the Identity Theory*

One way to characterize the history of modern philosophy is to recount the penchant of philosophical speculation to spawn empirical sciences, which, as they mature, return to commandeer intellectual domains on which philosophy had previously presumed to possess a proprietary claim. The growth of modern science over the past four centuries has been marked by groups of researchers explicitly adopting new terms ("physics," "chemistry," "biology," "psychology," "sociology," etc.) for designating the specialized inquiries that have resulted and for distinguishing those sub-fields from the whole of natural philosophy -- a term which has, not coincidentally, fallen (except in historical discussions) into total disuse.

We ask philosophical proposals for greater precision and detail, and in that process of pressing their conceptual resources, we expect them to organize, illuminate, and concur with our new discoveries about the world. What the birth of modern science brought were means for meeting such demands that are far more systematic, efficient, and penetrating than any devised before. The collective accomplishments of communities of scientific experts fostering theoretical competition, discovering empirical evidence, and monitoring the credibility of that evidence have proved far more effective at producing fruitful accounts of the world than isolated philosophical speculations where assessments usually rely on little more than ordinary language, common sense, intuitions, available anecdotes, thought experiments, and the canons of logic. Scientific standards encompass these considerations (at least as long as the common sense, the intuitions, and the anecdotes can withstand the critical scrutiny and progressive theorizing) as well as the far more exacting demand that theories meet and pass empirical tests, which scientists develop (using increasingly sophisticated experimental techniques) and pursue.

Naturalism in philosophy demands that philosophical proposals exhibit a healthy respect for the methods and findings of the empirical sciences, especially when they address the same domains that those sciences do. In the twentieth century philosophers became a good deal more circumspect about their physical and biological speculations. Science had become a *fundamental constraint* on credible metaphysical proposals about those domains.

The number of domains where philosophers must heed scientific developments has only increased as modern science has advanced. At the outset of the twenty-first century, philosophers who pronounce about matters of mind and language without regard to the cognitive sciences do so at their peril. When scientific research generates innovative schemes that are empirically testable, that systematically organize the pertinent phenomena, and that supply new explanatory and predictive insights, philosophers' declarations about what is imaginable or about what our concepts demand often appear quaint in retrospect. The pronouncements of contemporary philosophers of mind about what it must be like to have mental lives like ours or about unbridgeable gaps in scientific accounts of consciousness risk comparisons with Hegel's attempt to prove that there were only seven planets (Inwood 2003, p. 21).

Naturalists hold that *philosophy enjoys no privilege*. Typically, philosophers' only advantages arise from their wider views of things and their increased sensitivities to the structures and strengths of arguments. Certainly, philosophers' guesses are as good as anyone's. The suggestiveness of their guesses, however, does not obviate in the least the advantages that accrue to philosophical proposals that manifest familiarity with the sciences.

In less guarded moments some naturalists' enthusiasms about scientific progress have enticed them into entertaining the possibility of completely eliminating normative epistemology and the metaphysics that presently facilitates it (e.g., Churchland 1979, chapter five). There are two problems here.

First, the metaphysics behind presumptions about mental attitudes toward contents that informs normative epistemology substantially overlaps, at least currently, with conceptual commitments of the psychological and socio-cultural sciences. Consequently, this especially fervent version of naturalism generates a paradox, since fulfilling its goals would *appear* -- exclusively on the basis of its *philosophical projections* -- to jeopardize the status of entire sciences that have been up and running now for more than a century. (See McCauley 1986, 1996, and 2007.) This is paradoxical to the extent that all versions of naturalism aim, instead, to foster scientific initiatives and to restrain philosophical hubris.

The second problem with such fervent forms of naturalism is their failure to recognize that because the current conceptual framework in terms of which normative issues are formulated may not persist in the face of scientific progress in the cognitive and psychological sciences, it does not follow that the underlying normative concerns will disappear with them. (Paul Churchland's account (1989, p. 223) of "a virtuous mode of explanatory understanding" in terms of parallel distributed processing models of cognition signal growing moderation in his own version of philosophical naturalism about our interests in normative epistemology.) The sciences are usually quiet about the norms that pervade them and their associated practices. If naturalism is to include a robust picture of the scientific enterprise, then those norms are not just legitimate, but obligatory, targets for philosophical reflection. Although naturalists insist, contrary to traditional epistemology, that the sciences should constrain the categories from which we should expect to fashion our most compelling metaphysical and epistemological frameworks, we can never create those frameworks by simply doing more science. Getting better theories about the facts alone will not make those implicit norms explicit. Naturalism is not scientism. Its goal is not to put philosophy out of business. Philosophy still has plenty of jobs.

Questions remain, though, about how those jobs are best done. In the broadly transcendental tradition, philosophers such as Husserl (1970) and Thomas Nagel (1986) hold that some philosophical tool or insight provides philosophy with a unique form of analytical leverage with which it can explore such things as the very possibility of doing science. Other philosophers (e.g., Searle 1992) eschew the trappings of transcendental perspectives in favor of ordinary language and common sense (and even lay claim to a naturalistic orientation) but, nonetheless, pronounce no less confidently about the ways some things *must* be, either because our current concepts say so or -- what is nearly the same thing -- because common sense clearly shows that some scientific reductions are unthinkable. (See Churchland and Churchland 1998: chapters 8 and 9.)

More often than not, in the last century the privileged expectations under debate have concerned our *inner* natures, i.e., our mental lives, rather than the external world. These include everything *from* traditional phenomenology's presumptions about pure, mental exercises gaining

access to the contents and character of the mental representation undistorted by any theoretical commitments, to Nagel (1974) drawing epistemological conclusions about the character and limits of objectivity on the basis of what he takes to be inescapable presuppositions about human subjectivity, to proposed reductions of consciousness bemusing Searle, because he finds the proposed psycho-neural identities so obviously implausible on what are, basically, common sense grounds. (Searle underestimates just how counter-intuitive scientific achievements can be. See Churchland and Churchland 1998, p. 128 and McCauley 2000 and forthcoming, chapter three.)

For the naturalist, traditional philosophical tools and insights and attention to things like ordinary language and common sense are perfectly legitimate means for initiating inquiry and valuable propaedeutics to the formulation of more systematic, empirically accountable theories. The sheer inertia that many of these tools enjoy on the basis of their widespread appeal, their intuitive charm, and their long-standing philosophical service indicates that their counsel and influence should not be discounted unless it is fairly clear how each of those apparent virtues can be explained away (on a case by case basis).

Even if they cannot be explained away, though, for naturalists these considerations neither guarantee anything nor are they the whole story. These standard philosophical tools neither supersede nor diminish our obligations as inquirers to press our theories--as rigorously as we can--for greater precision, for greater detail, and for a continuing ability to make sense of new features of the world (such as findings about the consequences of various brain abnormalities and injuries). Why should simply sifting through the intuitions -- even the intuitions of particularly thoughtful, intelligent people -- that dominate at a particular time and place and checking them against the deliverances of a project in armchair sociolinguistics exhaust the methods of philosophy?

The history of science has regularly been a history of achieving what was once the unthinkable, the prevailing conceptual commitments to the contrary notwithstanding. The point of this section is that with the development and growing integration of experimental psychology, the cognitive sciences, and the neurosciences (especially cognitive neuroscience) and with the new tools for studying the activities of brains in connection with various tasks, naturalists not only have reasons to insist that these traditional philosophical methods do not provide the whole story about the connections between the physical and the mental, they have reasons to hold that those philosophical tools no longer even furnish the most important part of the story. There's the rub. My contention is that the principal obligation of contemporary philosophical proposals concerned with the relations of minds and brains is to accord with the best theories and findings of the pertinent sciences. That, at least, is HIT's aspiration. Assessments of the identity theory will turn primarily on the state of the relevant sciences and on deploying tools from the philosophy of science, especially those concerning cross-scientific relations.

3. *Heuristic Identity Theory*

Years ago in his reflections on the relations of science and epistemology, Willard Van Orman Quine (1969, p. 75) recommended that philosophy eschew make-believe. HIT does so by abandoning the philosophers' conceit that any positive verdict about psychoneural identity claims would only come after philosophers' prolonged collection and evaluation of evidence from our uses of ordinary language, our intuitions about our mental lives, and our imaginative exercises about possible worlds. The philosophies of psychology and neuroscience can no

longer afford to prize philosophical cleverness or metaphysical comfort over empirical accountability and explanatory adequacy.

The fact that it is counted as a truism amongst so many philosophers of mind that no empirical evidence could decide such matters only bolsters that conceit concerning the metaphysical character and the logical circumstances of psychoneural identity claims. Jaegwon Kim (1966, p. 227) captures the underlying presumption quite precisely: “. . . the factual content of the identity statement is exhausted by the corresponding correlation statement. . . . There is no conceivable observation that would confirm or refute the identity but not the associated correlation.” More recently, David Chalmers (1996, p. 115) puts the point as follows:

Neurobiological approaches to consciousness . . . can . . .
tell us something about the brain processes that are *correlated* with
consciousness. But none of these accounts explains the
correlation: we are not told why brain processes should give rise
to experience at all. From the point of view of neuroscience, the
correlation is simply a brute fact.

. . . Because these theories gain their purchase by *assuming*
a link . . . it is clear that they do nothing to explain that link.

The contention is that any evidence that can be cited to support an explanation of an identity is also perfectly consistent with affirming no more than correlations between psychological and neural phenomena. (As a principle guiding metaphysical deliberations, this truism seems to point in the opposite direction of Ockham’s famous razor.)

The problem, however, is that this deflationary view about the import of *any* empirical evidence for a psychoneural identity constitutes a misleading characterization of the place of hypothetical identities in *scientific* inquiries. HIT highlights two considerations bearing on the place of such hypothetical, cross-scientific identities in scientific research.

First, HIT stresses that, just as important as their standing as hard-won conclusions in a well-developed scientific research program, hypothetical identities also *regularly* serve as the critical *premises* in explanatory proposals that *inaugurate* new lines of scientific investigation. Hypothetical identities provide the logical and substantive leverage for motivating forays down completely new avenues of research. From Benjamin Franklin’s pursuit of evidence that lightning is, indeed, an electro-static discharge to the series of hypotheses about the location of humans’ visual cortex from the late nineteenth century through the late twentieth century (Bechtel and McCauley 1999; Bechtel 2008), hypothetical identities *initiate* new lines of research and point to new ways of obtaining evidence about the phenomena, the patterns, the systems, and the mechanisms under scrutiny.

Cross-scientific, hypothetical identities undergird an explanatory pluralism that showcases the enhanced theoretical, experimental, and evidential resources available to scientists. When scientists suggest identities that span levels of explanation -- say, a hypothetical identity between operations in the brain and some psychological function, such as describing some area in visual cortex as responsible for detecting colors -- they provide bridges for investigators at both levels of analysis. Those bridges enable researchers working at one analytical level to import theoretical ideas, experimental tools, and bodies of evidence from the other analytical level. Psychologists’ findings about the conditions under which people or animals do or do not, in fact, detect the critical features in question direct neuroscientists’ designs of experiments when observing the activities of brains. Neuroscientists’ findings about

areas of the brain that are active when performing some task suggest to psychologists, by virtue of their command of earlier, related findings in the psychological literature, other tasks that should be systematically related to the focal task (whether by association or dissociation).

In sum, then, the first difference between HIT's take on psychoneural identities and that of conventional philosophy of mind is that hypothetical identities are not (only) the conclusions emerging from decades of research in cognitive neuroscience. They are just as, if not more, important as *heuristics of scientific discovery* in the first stages of research.

It is in just this respect that conventional philosophical treatments of the multiple realizability of psychological states, whether across functionally equivalent systems composed of different materials or across species or across individual members of some species or across the same individual at different times, prove a misleading account of the dynamics of cognitive neuroscience.

The first step is to domesticate multiple realization by pointing out how often it arises in nature and how it does *not* forestall cross-scientific, hypothetical identities at other levels of analysis in science. The Churchlands (1998, p. 78), for example, note that "in a gas, temperature is one thing; in a solid, temperature is another thing; in a plasma, it is a third; in a vacuum, a fourth; and so on. . . . this . . . just teaches us that there is more than one way in which energy can be manifested at the microphysical level." As Robert Richardson (1979; 1982) has emphasized, reductions in science are *domain specific*. For many purposes the division of psychology and cognitive science into specialized sub-domains seems plausibly motivated on a variety of criteria (in the same way that accounts of heat in gases, solids, plasmas, vacuums, and so on are usefully distinguished for some of our problem solving purposes in physical science). (Mundale and Bechtel 1996, p. 490)

Multiple realizability arguments pertaining to human brains look plausible, first, because anti-reductionist philosophers have generally failed to attend to what scientists have ascertained to be the theoretically significant kinds at each analytical level (especially at the level of neuroscience) and, second, because they have ignored whether the kinds they do discuss are cast at comparable grains. With regard to the first consideration, science is about ascertaining which resemblances and differences *matter* from the standpoints of explanation, prediction, and control.¹ The aim is not to map each and every homespun category we may employ, but rather to concentrate on those that our best explanatory theories spotlight (Hardcastle 1996). With regard to the second consideration, philosophers find ubiquitous multiple realizability in psychology because they regularly compare coarse grained psychological concepts with exceedingly fine-grained conceptions of brain states. The folk psychological notions that particularly interest philosophers are more coarse grained than most employed in experimental cognitive psychology,

¹ Such considerations may even neutralize the sting of Jerry Fodor's (1974) famous argument about the fruitlessness -- for understanding economics -- of a focus on the various instantiations of money. Attention to the limitations that particular material forms that money can take impose on transactions will disclose some eminently useful, though admittedly low level, generalizations about those forms' deployment within economies. For example, some transactions such as mortgage closings at banks and purchases of items stored in the inside pockets of less scrupulous vendors' trench coats in alleys in large cities will almost never involve personal checks or credit cards or, at least at the mortgage closings, large amounts of cash. Thus, some patterns in the economic domain may offer grounds for the fragmentation of the concept 'money' along these lines for certain limited, domain specific explanatory purposes.

while the conceptions of brain states they discuss, Bechtel and Jennifer Mundale argue, are much finer-grained than the ones practicing neuroscientists use in their theories. They comment (1999, p. 178) that “when a common grain size is insisted on, as it is in scientific practice, the plausibility of multiple realizability evaporates.” Ascertaining compatible grains between inquiries at two different levels fosters the co-evolution of sciences. Getting the grains right between theoretically significant kinds can make all the difference. A variety of *successful research strategies* from the border areas between psychology and neuroscience, some of which have, by now, been *utilized for more than a century*, indirectly repudiates the multiple realization of theoretically relevant psychological states.

This is not only true about the interpretation and the integration of recent findings from PET (positron emission tomography) and fMRI (functional magnetic resonance imaging) research but also about much older inferences that neuroscientists have made about the cognitive functions of various areas in unimpaired brains on the basis of studies of performance deficits and brain damage. Incorporating these considerations does not simplify the story, but their links to scientific theorizing and empirical findings in the physical and biological sciences, with regard to the imaging cases, and in the biological and behavioral sciences with regard to deficits and damage, do help to solidify the story.

For example, consider PET imaging. PET imaging involves multiple assumptions about a host of physical and biological processes including: (1) that heightened neural activity consumes energy, (2) that the energy is derived from reactions among chemicals from the blood, (3) that demand for increased energy requires the delivery of greater amounts of blood, (4) that detecting heightened blood flow in some area could be detected by injecting water marked with O15 (an isotope of oxygen that is radioactive) into the blood, (5) that the marked water molecules will release a positron, usually sooner rather than later, (6) that some nearby electron and the positron will annihilate one another, (7) that their mutual annihilation will yield a gamma ray with a characteristic wave length, (8) that the gamma ray will pass through the matter in humans’ heads, and (9) that detectors of the proper construction will detect those gamma rays.

Or, as a second example, consider the fact that much research with the various neuroimaging technologies employs the subtraction method. It assumes that the differences between the neural activities associated with two tasks, where one includes every aspect of the other plus some further process of interest, will furnish information about the neural activity associated with that process of interest. (See Roskies 2010 and Van Orden and Paap 1997.) Crucially, the subtraction method also assumes that the brain exhibits at least some stable “functional-anatomical” specialization “over time and across populations” (Roskies 2010, p. 654).

The point in both of these examples is that neither the pervasiveness, nor the variety, nor the detail of such assumptions undermines the use of these technologies. PET, fMRI, and other imaging studies, employing the subtraction method, regularly disclose significant differences in the levels of activation or in the pattern of areas that are activated or both.

Bechtel and I (1999) also push the case against multiple realizability beyond our species by stressing the importance of recalling that, until recently, most research in neuroscience was done on the brains of non-human animals. Identifications of brain areas and processes were done *comparatively*. The multiple realization of some psychological function across species in *homologous* structures did not obstruct the identification of some function with an area. On the

contrary, it was *one of the most compelling types of evidence* available for identifying an area in the human brain and assigning it a function. Contrary to contemporary anti-reductionist orthodoxy, multiple realizations across species are not a barrier to the mapping of some psychological function on to brains. Historically, they were one of the keys to accomplishing such mappings.

This leads straightaway to the second important difference between HIT's account of cross-scientific identities between psychological and neuroscientific models and most of the prevailing philosophical accounts. Although HIT acknowledges that some cross-scientific identities *are* fairly understood as the hard won (yet provisional) conclusions of extended scientific investigations, they merit that status on very different grounds than the ones on which most philosophers of mind seem to imagine. (Recall Chalmers' demand for an *explanation of the link* between neural processes and conscious mentality.)

Cognitive neuroscientists show why some neural mechanism might *constitute* some psychological phenomenon by exploring the empirical success of the wide range of predictions and explanatory connections that assumption generates. *It is that empirical success that corroborates the constitutive hypothesis and tentatively justifies its assumption* (Churchland and Churchland 1998, pp. 120-122). The tentativeness of the justification here is nothing special. It is the same tentativeness about justification that accompanies every scientific claim, and it is that tentativeness that informs the hedge in the previous paragraph about so-called "hard won" conclusions.

HIT directly challenges Chalmers' claim that "[f]rom the point of view of neuroscience, the correlation is simply a brute fact." It is precisely from the perspective of neuroscience (*and* from the perspective of psychological science) that the correlated phenomena are *explained by the hypothetical identity*. (See Hill 1991, pp. 22-26.) That identity is, in turn, warranted by the explanatory and predictive successes that it informs and by the productive program of empirical research it inspires. HIT underscores the fact that evaluations of proposed identities do not turn on confirming them directly. What, after all, could *that* possibly be? (McCauley 1981) The evidence for an identity claim arises *indirectly* -- primarily on the basis of the emerging empirical successes of the explanatory hypotheses it motivates. For example, if normal activities in V4 are identical with the processing of information about wave length, then serious abnormalities of particular types in the structure and functioning of V4 should yield abnormalities of particular types in subjects' color perception. The point is that this hypothetical identity is an empirical conjecture that researchers can use both psychological and neuroscientific evidence not only to assess but to *refine*.

Obtaining indirect corroborating evidence for identifying some neural process with some psychological function along such lines no more finalizes that identity than it would any other hypothesis in science. Nor does it establish that the function under scrutiny is either the sole or even the primary function these neural processes realize. Still, the more hypotheses of this sort the identity informs and the more successful those hypotheses prove, the more likely the identity will come to serve as an assumption the sciences lean upon rather than a bare conjecture in search of support (Van Gulick 1997). Such identity claims are, of course, no less conjectures still. They are, however, no longer simply *bare* conjectures (let alone "brute facts")!

HIT *shifts the grounds for the debates* about the plausibility and merits of the psycho-physical identity theory. Of a piece with the naturalists' general agenda, HIT certainly proposes to move the assessment of the identity theory *beyond* the rarefied domain of philosophical

reflection on the conceptual limits of the imaginable and on intuitions (whether about our own mental lives or about deeper metaphysical matters) to the rough and tumble, on-going activities of the psychological and neural sciences. Naturalists contend that the satisfactoriness of any version of the identity theory should be assessed *primarily* on the basis of its ability to make sense of the prevailing theories, findings, and activities of the relevant psychological and neural sciences.

Most contemporary philosophers find it unexceptional that metaphysical proposals about the nature of life or about species or about phylogeny or about other biological topics should stand on all fours with our best theories and practices in the biological sciences. Nor do they question that the failure of metaphysical proposals to do so should count *prominently* against such proposals. HIT maintains that a parallel moral, with respect to the psychological and neural sciences and the matters they address, applies to proposals in contemporary philosophy of mind. The cognitive, psychological, and neural sciences have all reached the age of majority.

4. *Localizing Human Face Perception in the Brain*

The study over the past two decades in psychology and neuroscience of the human capacity to recognize human faces furnishes something of a parade case of the cross-scientific dynamics in action that HIT emphasizes. A sketch of some of the most prominent interactions must suffice in what follows, but I must skate over many details in the interest of space limitations, not because any of those details are any less likely candidates for this HIT parade.

The familiar comment that “I never forget a face” is but one common manifestation of most human beings’ confidence that they have elevated levels of memory for human faces, as compared, say, to their memories for the labels on wine bottles or, more notoriously, compared to their memories for people’s names. That widespread impression is at least partly a function of the fact that, in most social circumstances, the two memory tasks are disparate, since we typically must *recall* people’s names but we only have to *recognize* their faces. Various controlled studies on that front (e.g., Faw 1990) have yielded equivocal results, but another possibility is that humans may have better performance on recognition memory for both faces and names compared to other sorts of items.

Some theorists (e.g., Tooby and Cosmides 1992, p. 97) in evolutionary psychology have proposed that humans have an evolved, domain-specific face recognition mechanism. The scenario is straightforward. The detection and identification of individual conspecifics carries vital importance for any animal, but especially for the hyper-social species, *Homo sapiens*, for which faces hold keys not merely to individual identities but to organisms’ emotional and mental states. This seems to be an example, however, for which the evolutionary psychologists do not have to rely on an adaptationist scenario to do all of the heavy lifting (cf. Richardson 2007).

Independently, in both psychology and neuroscience in both experimental and clinical pursuits, researchers have explored the peculiarities and patterns behind face recognition and the brain areas and the connections that seem to be most prominently involved in these and related functions. Nancy Kanwisher, Josh McDermott, and Marvin Chun (1997, p. 4302) open their landmark paper on the Fusiform Face Area (FFA) listing the many disciplines (cognitive psychology, computational modeling of vision, neuropsychology, and neurophysiology) and some of the tools (single cell recordings in both animals and humans and selective deficit studies) that had led researchers to consider the possibility of a functionally isolable, neural module for face perception. They further highlight a variety of studies that had used fMRI to

ascertain brain areas that demonstrated greater activity in response to viewing faces as opposed to objects or scrambled faces or consonant strings and in response to matching faces as opposed to locations. Since such findings are consistent with a variety of plausible hypotheses (other than the proposal that these areas are peculiarly active in face perception), Kanwisher et al. (1997) use fMRI to look responses to a collection of alternative stimuli aimed at ruling out those alternative hypotheses.

On the basis of considerable previous research, Kanwisher et al. (1997) focused on occipitotemporal areas in the ventral pathway that, in the passive viewing of pictures of faces compared to the passive viewing of pictures of objects, responded with significantly greater activity. They found that for all ten of their right-handed participants the right fusiform gyrus proved significantly more active with face stimuli as opposed to object stimuli. Half of that group as well as one of their left-handed subjects showed bilateral activation, involving the left fusiform gyrus as well. Their other left-handed subject showed unilateral activation in the left fusiform gyrus only. Kanwisher et al. 1997 (p. 4306) comment that “[d]espite some variability, the locus of this fusiform face activation is quite consistent across subjects both in terms of gyral /sulcal landmarks and in terms of Talairach coordinates.” In experiments testing pictures of faces against pictures of houses, of three quarter views of faces (with hair under ski hats), and of human hands, Kanwisher et al. (1997) demonstrate that the fusiform gyrus reacts with significantly greater activity to the pictures of faces, as it also does to pictures of faces in a task requiring participants to detect matches between the current stimulus and the one immediately before.

Certainly, the relevant literature they review and the experimental evidence they provide are suggestive, but Kanwisher et al. (1997, p. 4309, emphasis added) maintain that “[t]he elimination of these main alternative hypotheses provides *compelling evidence* that the fusiform face area described in this study, which we will call area “FF,” is specifically involved in the perception of faces.” After providing additional negative experimental evidence against a few more of the prominent alternative hypotheses, Kanwisher, McDermott, and Chun identify activity in the FFA with the (psychological) task of face perception. Their protestation about “compelling evidence” notwithstanding, they have advanced a theoretically provocative, cross-scientific, *hypothetical identity*.

HIT holds that such a proposal will provoke research aimed at testing and refining this hypothetical identity. That observation immediately raises questions about the grain of the items being identified. On the basis of their imaging studies, Kanwisher et al. (1997) have narrowed things down considerably at the neural level. Rather than looking at the entire occipitotemporal cortex, their studies isolate the FFA as the consistently active structure in the wide variety of face perception tasks they posed for their participants. Kanwisher et al. (1997) demonstrated a double dissociation with regard to pictures of faces versus pictures of objects for the FFA, on the one hand, and for “a different, bilateral and more medial area” as well as for the parahippocampal region, on the other (p. 4304).

Comparatively speaking, the grain at the psychological level is coarser. It is also left somewhat vague. Kanwisher et al. (1997) comment that “[f]or present purposes, we define face perception broadly to include any higher-level visual processing of faces from the detection of a face as a face to the extraction from a face of any information about the individual’s identity, gaze direction, mood, sex, etc.” From a functional standpoint, these constitute a diverse list of features that pertain to a wide variety of possible psychological systems from theory of mind, to

the detection of emotions, to the detection of kin or possible mates, to a person file system. Finally, though, Kanwisher, McDermott, and Chun narrow the psychological grain and, thus, sharpen the hypothetical identity themselves (the ambiguity of the phrase “specifically involved” notwithstanding). They propose (p. 4310) that “the human fusiform area is specifically involved in the discrimination of individual identity.” Of course, their proposal is not idle speculation. Consonant with HIT’s explanatory pluralism, they offer this conjecture on the basis of what appear to be homologies with areas that earlier research about face recognition capacities in macaques have suggested are relevant.

They explicitly underscore (p. 4310) that their research makes strides toward resolving the problem of achieving comparable grains between the neural and the psychological: “[o]ur use of a functional definition of area FF allowed us to assess the variability in the locus of the “same” cortical area across different individual subjects.” They also accentuate the fact that not all of their subjects exhibited bilateral activation in the fusiform gyrus, consistent with the growing consensus among researchers that damage to the right side only does not always produce prosopagnosia, i.e., the inability to identify individual humans by recognizing their faces, regardless of being able to see normally otherwise.

It is worth noting how untroubled Kanwisher and her colleagues are about some minor variability in the location of the FFA across subjects. Most of the findings they report, as in the vast majority of papers that use the new imaging technologies to study neural activities across a group of participants, are *averages* across the population of brains they are studying. Multiple realizations across individuals appear to trouble contemporary cognitive neuroscientists no more than multiple realizations across species troubled early cognitive neuroscientists. Kanwisher et al. (1997, p. 4310) found and used similar levels of activity on face perception tasks in what were basically the same areas across subjects to locate the FFA, i.e., “the fusiform gyrus or the immediately adjacent cortical areas in most right-handed subjects.” They report that this locus of activation is very similar to those spotlighted in other research on face processing and “virtually identical in Talairach coordinates to the locus reported in one (40x, -55y, -10z for the mean of our right-hemisphere activations; 37x, -55y, -10z in Clark et al., 1996).”

Kanwisher et al. (1997, p. 4310) comment that their proposal about a localized portion of neural machinery, which specializes on face perception, counts against “a single general and overarching theory of visual recognition.” Subsequently, in a collection of further papers Kanwisher and her various colleagues carry out studies aimed at advancing both the identification of heightened levels of neural activation in the FFA with face perception and face identification, in particular, and the related thesis that face perception is a domain specific capacity to be distinguished from the perception of other kinds of things. A representative group of those papers supply evidence for such things as: (1) that FFA activity is not associated with either the development or exercise of just any expertise but concerns the processing and identification of *human* faces (Kanwisher 2000 and McKone, Kanwisher, and Duchaine 2006), (2) that both particular facial features and global facial configurations elicit FFA activation, though memories of individual faces for the purposes of face identification may be stored in higher-level areas (Tong et al. 2000), (3) that responses occurring in occipitotemporal cortex at both 100ms and 170ms after stimulus onset seem to be correlated with recognition that a stimulus is a face (face categorization) but that only the second response at 170ms is correlated with face identification (Liu, Harris, and Kanwisher (2002)), and (4) that information about the parts of

faces and information about their spacing are processed by specialized, holistic mechanisms (Yovel and Kanwisher 2008).

Among the wide variety of dissent that the hypothetical identity of heightened FFA activity and face perception has inspired, I will briefly discuss but two papers (Hasson et al. 2003; Steeves et al. 2006). The overall logic of both is the same. Both compared the processing of various stimuli, including pictures of faces, in the brain of a prosopagnosic participant with processing in the brains of control participants and found evidence that heightened FFA activation was *not sufficient* for the *identification* of faces. Exploring possible consequences of the hypothetical identity of Kanwisher et al. (1997), they have pursued research that suggests the need to achieve an even finer grain at the psychological level, resulting in the hypothetical identity's emendation and refinement. Both suggest that heightened FFA response to faces is concerned with the *detection* of faces instead of their identification.

Uri Hasson and his colleagues studied a congenital prosopagnosic participant, Y.T., who was an otherwise healthy, thirty-nine year old businessman. Although YT had neither any anatomical lesions nor any history of neurological disease, he had had a severe deficit in identifying people's faces for as long as he could remember. Despite his inability to identify individual faces, YT could readily ascertain gender, age, and emotion from human faces. Hasson and his colleagues used standard means with fMRI for determining the areas of YT and control participants' brains that were most active when dealing with pictures of faces. Those areas corresponded to findings from earlier research and included both an area in the lateral occipital cortex (LO) and *the FFA in all of their participants, including YT* (Hasson et al. 2003, p. 422). Unlike people with acquired prosopagnosia, congenital prosopagnosics (YT, at least) can exhibit the same levels of activation in the same anatomical locations with the same hemispheric laterality as normal participants in response to pictures of faces. With but one exception, viz., the left LO, YT's activation levels on the face perception tasks were within one standard deviation of the control group's means and even the difference in the left LO did not quite reach significance ($p < .053$). (Hasson et al. 2003, p. 422)

YT's response at 170ms, as revealed by evoked response potentials (ERP), however, showed a comparable LO response for objects as for faces. (Hasson and his colleagues point out that Sagiv et al. 2000 also found this pattern in two other congenital prosopagnosic participants.) In their second experiment Hasson and his colleagues (in different experimental conditions) foregrounded in the Rubin Face-Vase stimulus either the faces on the sides or the vase in the center (by using a bright solid color in the fore-grounded region and vertical lines in the other). Their findings counted against the hypothesis that YT's heightened levels of activity in response to faces was by virtue of his FFA only being able to handle face parts (as it also handled objects), and, thus, being unable to integrate them. YT's responses fell within one standard deviation from the controls' mean for each category (i.e., faces versus vases), indicating that activity in YT's FFA and LO were driven, in part, by the *holistic* arrangement of faces.

Hasson et al. (2003, p. 426) argue that the fact that YT's FFA seems to process faces in the same ways that controls' brains do suggests that heightened FFA activation is not sufficient for identifying individuals' faces but only for *detecting* human faces. (*Prima facie* this would also seem to square with YT's indiscriminate LO response at 170ms.) That YT can distinguish faces from non-faces suggests this, as does his abilities to discriminate facial expressions and gender. That FFA activation constitutes face detection would indicate that its heightened activation is a normal and necessary step in face identification, but the constellation of findings

concerning YT argues that it is not sufficient for carrying out that task. That claim would still be consistent with findings that *acquired* prosopagnosics often manifest a lack of FFA activation for one reason or another. (Steeves et al. 2006, p. 605)

On the basis of findings about an acquired prosopagnosic (versus controls), Jennifer Steeves and her colleagues (2006), in fact, argue for the *same* allocation of functions as Hasson and his colleagues. DF is a forty-seven year old female, whose brain was damaged by accidental carbon monoxide poisoning when she was thirty-four. She clearly has a severe deficit with regard to object recognition. Only later was her prosopagnosia discovered and studied. DF shows no higher level face processing. She “. . . recognizes people . . . on the basis of non-face cues such as clothing, hair, stature, gait and voice . . . D.F. does not respond to facial expression in her interaction with others” (Steeves et al. 2006, p. 596). DF completely failed a recognition test of famous people, and she was unable to determine faces’ gender reliably. DF is unable to differentiate normal and scrambled faces when they are presented sideways, and only with considerable time can she distinguish upright from inverted faces but not if they are partially occluded on the vertical midline (Steeves et al. 2006, pp. 601-603). These findings suggest how she can, nonetheless, reliably discriminate normal, upright faces from objects by attending to their vertical symmetry and their overall face-like configurations.

The crucial point is that in response to faces DF shows both *the same location and “similar” activation for FFA* that is “comparable” to that in at least one of their control participants (Steeves et al. 2006, pp. 597 and 606). DF suffers from bilateral lesions in the occipital face area (OFA) in LO, which is a “relatively early visual area in the ventral stream” (Steeves et al. 2006, p. 595). Steeves et al. 2006 (pp. 606-607) propose that higher level face processing and face identification require “intact connections with the OFA area” via “feedback connections from the FFA to the OFA.” With an intact FFA recognizing the overall configuration (of upright faces), DF retains the ability to categorize, i.e., detect faces, but her ability to carry out higher level facial processing and face identification are undone

. Thus, Steeves et al. 2006, looking at the profile of an acquired prosopagnosic, DF, and Hasson et al. 2003, looking at a congenital prosopagnosic, YT, both argue that normal processing in FFA is not sufficient for face identification, though it is probably necessary. That accords with their common contention that what FFA does is face detection.

As HIT maintains, the hypothetical identity of Kanwisher et al. 1997 between face identification and FFA activity motivates a subsequent wave of research that tacks back and forth between psychological findings about human performance and neuroscientific findings about brain activity. This results not only in further insight and more nuanced theory about face processing in normal participants as well as in congenital and acquired prosopagnosics but in an even more finely honed hypothetical identity of psychological function and neural processing.

* I wish to express my gratitude to Simone Gozzano and Christopher Hill for their helpful comments on an earlier draft of this paper.

References

- Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Routledge.
- Bechtel, W. and McCauley, R. N. (1999). Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem against the Background of Research Strategies in Cognitive Neuroscience. In *Proceedings of the twenty-first meeting of the Cognitive Science Society*. M. Hahn and S. C. Stones (eds.). Mahway, New Jersey: Lawrence Erlbaum Associates, 67-72.
- Bechtel, W. and Mundale, J. (1999). Multiple Realizability Revisited: Linking Cognitive and Neural States, *Philosophy of Science* 66, 175-207.
- Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G., Haxby, J. V. (1996) Functional Magnetic Resonance Imaging of Human Visual Cortex During Face Matching: A Comparison with Positron Emission Tomography, *NeuroImage* 4, 1-15.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Churchland, P. M. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Churchland, P. M. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: The MIT Press.
- Churchland, P. M. and Churchland, P. S. (1998). *On the Contrary*. Cambridge, MA: MIT Press, 65-79.
- Faw, H. (1990). Memory for Names and Faces: A Fair Comparison, *American Journal of Psychology*, 103, 317-326.
- Fodor, J. A. (1974). Special Sciences (or: The Disunity of Science as a Working Hypothesis), *Synthese* 28, 97-115.
- Hardcastle, V. G. (1996). *How to Build a Theory in Cognitive Science*. Albany: SUNY Press.
- Hasson, U., Avidan, G., Deouell, L. Y., Bentin, S., and Malach, R. (2003). Face-selective Activation in a Congenital Prosopagnosic Subject, *Journal of Cognitive Neuroscience* 15, 419-431.
- Hill, C. S. (1991). *Sensations: A Defense of Type Materialism*. New York: Cambridge University Press.

- Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. D. Carr (trans.) Evanston: Northwestern University Press.
- Inwood, M. (2003). *A Hegel Dictionary*. Oxford: Blackwell Publishers.
- Kanwisher, N. (2000). Domain Specificity in Face Perception, *Nature Neuroscience* 3 (8), 759-763.
- Kanwisher, N., McDermott, J., Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception, *Journal of Neuroscience* 17, 4302-4311.
- Kim, J. (1966). On the Psycho-physical Identity Theory. In *Materialism and the Mind-Body Problem*. D. Rosenthal (ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 80-95.
- Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of Processing in Face Perception: An MEG Study, *Nature Neuroscience* 5, pp. 910-916
- McCauley, R. N. (1981). Hypothetical Identities and Ontological Economizing: Comments on Causey's Program for the Unity of Science, *Philosophy of Science* 48, 218-27.
- McCauley, R. N. (1986). Intertheoretic Relations and the Future of Psychology, *Philosophy of Science* 53, 179-99.
- McCauley, R. N. (1996). Explanatory Pluralism and the Coevolution of Theories in Science, *The Churchlands and Their Critics*. R. N. McCauley (ed.). Oxford: Blackwell Publishers, 17-47.
- McCauley, R. N. (2000). The Naturalness of Religion and the Unnaturalness of Science. In *Explanation and Cognition*. F. Keil and R. Wilson (eds.). Cambridge: MIT Press, pp. 61-85.
- McCauley, R. N. (2007). Reduction: Models of Cross-Scientific Relations and Their Implications for the Psychology-Neuroscience Interface. In *Handbook of the Philosophy of Science: Philosophy of Psychology and Cognitive Science*. P. Thagard (ed.). Amsterdam: Elsevier, pp. 105-158.
- McCauley, R. N. (forthcoming). *The Naturalness of Religion and The Unnaturalness of Science*.
- McCauley, R. N. and Bechtel, W. (2001). Explanatory Pluralism and The Heuristic Identity Theory, *Theory and Psychology* 11, pp. 738-761.
- McKone, E., Kanwisher, N., and Duchaine, B. D. (2006). Can Generic Expertise Explain Special Processing for Faces? *Trends in Cognitive Science* 11, pp. 8-15.

- Mundale, J. and Bechtel, W. (1996). Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function, *Minds and Machines* 6, 481-505.
- Nagel, T. (1974). What is it Like to Be a Bat? *Philosophical Review* 83, 435-50.
- Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.
- Place, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*, 47, 44-50,
- Quine, W. V. O. (1969). Epistemology Naturalized, *Ontological Relativity and Other Essays*. New York: Columbia University Press. 67-90.
- Richardson, R. C. (1979). Functionalism and Reductionism, *Philosophy of Science* 46, 533-58.
- Richardson, R. C. (1982). How Not to Reduce a Functional Psychology, *Philosophy of Science* 49, 125-37.
- Richardson, R. (2007). *Evolutionary Psychology as Maladapted Psychology*. Cambridge: MIT Press.
- Roskies, A. L. (2010). Saving Subtraction: A Reply to Van Orden and Paap, *British Journal for the Philosophy of Science*, 63, 635-665.
- Sagiv, N., Barnes, L. L., & Robertson, L. C. (2000). The Neural Bases of Congenital Prosopagnosia, *Society for Neuroscience Abstracts*, 26.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Steeves, J.K.E., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., Humphrey, G.K., Milner, A.D., and Goodale, M. A. (2006). The Fusiform Face Area is Not Sufficient for Face Recognition: Evidence from a Patient with Dense Prosopagnosia and No Occipital Face Area, *Neuropsychologica* 44, pp. 594-609.
- Thagard, P. (2010). *The Brain and the Meaning of Life*. Princeton: Princeton University Press.
- Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., and Kanwisher, N. (2000). Response Properties of the Human Fusiform Face Area, *Cognitive Neuropsychology* 17, 257-279.
- Tooby, J. and Cosmides, L. (1992). The Psychological Foundations of Culture. In *The Adapted Mind*. J. Barkow, L. Cosmides, and J. Tooby (eds.). New York: Oxford University Press, 19-136.
- Van Gulick, R. (1997). Understanding the Phenomenal Mind: Are We All Just Armadillos? *The Nature of Consciousness: Philosophical Debates*. N. Block, O. Flanagan, and G. Güzeldere (eds.). Cambridge: MIT Press.

Van Orden, G. C. and Paap, K. R. (1997). Functional Neuroimages Fail to Discover Pieces of Mind in Parts of the Brain, *Philosophy of Science* 64, S85-S94.

Yovel, G. and Kanwisher, N. (2008). The Representation of Spacing and Part-Based Information Are Associated for Upright Faces But Dissociated for Objects: Evidence from Individual Differences, *Psychonomic Bulletin* 15, 933-939.