

Supporting Information

Kanwisher 10.1073/pnas.1005062107

SI Text

Brief History of the Debate About Localization of Function in the Brain. After his auspicious beginning, proposing that the brain is the seat of the mind and that the mind is composed of distinct mental faculties, Gall went on to argue that aptitude for a given faculty could be read from bumps on the skull. He further proposed 27 specific organs of the brain, including those for metaphysics, pride, the poetic talent, religion, the carnivorous instinct, and philoprogenitiveness (love of one's offspring). The new field of phrenology caught on, making appearances in the works of Melville and the Brontës (and, much later, Stephen Colbert). A popular book on phrenology published in 1828, *The Constitution of Man* by George Combe, sold over 100,000 copies, more than any other book in English in its time except the *Bible* and *The Pilgrim's Progress* (1).

Phrenology was controversial from the start. An early opponent was the French physiologist Jean Pierre Flourens (1794–1867), who saw the brain as a largely undifferentiated general-purpose organ. Among the first to lesion animal brains in the name of science, Flourens argued that “all sensations, all perceptions, and all volition occupy concurrently the same seat in [the brain]. The faculty of sensation, perception, and volition is then essentially one faculty” (1). Celebrities also joined the fray. Napoleon objected to Gall's materialism (2), and the Holy Roman Emperor Franz II issued a decree banning Gall's lectures on the grounds that Gall's system was discussed with excessive zeal, including by women, and that his system contradicted “the first principles of morality and religion” (3). This decree seems to have been the work of Franz II's jealous personal physician, and its main effect was to stoke public interest in Gall's views (3).

Fighting back, Gall's student Jean-Baptiste Bouillaud viciously attacked Flourens, commenting that “[i]f facts were not present in mass to fight [Flourens'] assertion, a minimal amount of reasoning would be enough to refute it” (4). It was Bouillaud who first made systematic use of evidence from brain-damaged patients, concluding that “[i]n the brain there are several special organs. . .” (1), and further concluding that one of these was for speech (2).*

The idea of functional localization entered the academic mainstream only when the highly respected Paul Broca announced to the Anthropological Society in Paris in 1861 that the left frontal lobe was the seat of speech. Broca's evidence came from a patient nicknamed Tan, because this was the only word he could say after damage to the inferior part of his left frontal lobe. Crucially, Broca noted that Tan's deficit was quite specific to speech; his other intellectual capacities remained intact. Although the debate about localization of function continued, a consensus emerged in the early 20th century: at least some basic sensory and motor functions reside in specialized brain regions.

The debate continued on the two questions of (i) how functionally specialized are regions of the brain and (ii) whether only basic sensory and motor functions are carried out in functionally specialized regions, or whether the same might be true even for higher-level cognitive functions.

Do Category-Selective Regions Contribute only to the Perception of Their Preferred Stimuli? The argument in the text that the fusiform face area (FFA), parahippocampal place area (PPA), and ex-

trastriate body area (EBA) are functionally specific rested heavily on the observation that each of these areas responds much more strongly to stimuli of its preferred class than to other stimuli. Each of these regions, however, does respond significantly (albeit weakly) to objects that are not in the preferred category (aka nonpreferred objects) (5). Further, in an important challenge to the claimed specificity of these regions, Haxby et al. (6) reported that the spatial pattern of response across the FFA contains information about nonfaces, that the pattern of response within the PPA contains information about nonscenes (3), and hence, that “[r]egions such as the ‘PPA’ and ‘FFA’ are not dedicated to representing only spatial arrangements or human faces, but, rather, are part of a more extended representation for all objects” (6).

Three important questions must be addressed here. First, do the FFA and PPA in fact contain information about nonpreferred stimuli? Second, even if these regions do contain some information about nonpreferred stimuli, is this information present under natural viewing conditions? And third, even if information about nonpreferred stimuli is present, including under natural viewing conditions, is it used in perception, or merely epiphenomenal? I address each question in turn.

To answer the first question, we can test for the presence of information in a given brain region (say, the FFA) by asking whether the spatial pattern of response across that region to one nonface object category is more similar and replicable across repeated measures (i.e., correlated) than that pattern produced by another object category.[†] Fancier machine-learning methods ask the same question by testing whether a variety of classifiers (for example, linear-support vector machines) can discriminate the categories on the basis of the pattern of response in the region under consideration. Although these methods at first produced inconsistent results (6–8), later studies have shown that a small but significant amount of information about nonpreferred stimuli indeed exists in the pattern of response within the FFA and within the PPA (9, 10).

An important caveat should be mentioned about the relation between the information detectable in patterns of fMRI response versus the information present in the actual neural code. On the one hand, fMRI can overestimate the role of category-selective regions in the representation of nonpreferred stimuli, because limits in the spatial resolution of fMRI virtually guarantee some blurring of the response of a target region with the response of its cortical neighbors. Consistent with this intuition, individual neurons in face-selective patches in monkeys (recorded electrophysiologically) show greater face selectivity than the same patches show with fMRI (11). On the other hand, any information we can see in the fMRI patterns is likely to be a small subset of the actual information present at the much finer grain of populations of spiking neurons. Thus, fMRI in some ways overestimates and in other ways probably underestimates the information present about nonpreferred stimuli. Still, current physiological evidence (11) is consistent with the evidence from

*Bouillaud was apparently the first to articulate the logic of the double dissociation—that two mental skills are embodied in separate brain areas if each can be disrupted in isolation from the other (4). Even today, this method remains one of the most powerful sources of information on the functional organization of the human brain.

[†]This claim can be visualized by first realizing that each region is a piece of the 2D cortical surface and then imagining the spatial pattern of response across that region as a hilly landscape in which altitude corresponds to response magnitude. The claim then that the FFA contains information about nonfaces amounts to saying that the shape of the landscape of response in the FFA is reliably different when people look at (say) cars versus when they look at (say) shoes, even if the mean altitude across the whole region is the same for cars and shoes.

fMRI that selective regions carry a small but significant amount of information about nonpreferred stimuli.

Second, analyses of the spatial pattern of the fMRI response within the ventral visual pathway have been based on the fMRI responses elicited by single cut-out objects on a blank background, presented at the fovea. Of course, real-world visual stimuli are not this simple: a typical visual scene contains multiple objects and complex background textures (what vision scientists call “clutter”). So, even if a small amount of information is available about nonpreferred objects in category-selective regions of cortex, is that information still present when subjects view cluttered displays more typical of real-world vision? Leila Reddy and I tested a simple version of this question, with two objects present simultaneously in the visual field (both on blank backgrounds). We found that when single cut-out stimuli were shown one at a time, the pattern of response in the FFA contained considerable information about faces and significant although weaker information about nonfaces. Similarly, the pattern of response in the PPA contained robust information about houses (which activate the PPA strongly, although not as strongly as a full scene) and significant but weak information about nonplaces. Crucially, however, when two objects were present at one time, information about preferred stimuli was virtually undiminished from the single-object case, but information about nonpreferred stimuli dropped to insignificance (9). This study and later related studies (12) suggest that category-selective regions may have little or no information about nonpreferred stimuli under more natural (i.e., cluttered) viewing conditions.

Still, given that fMRI is bound to underestimate the information present in the full neural population code, it is possible that future physiological studies will reveal some information about nonpreferred stimuli in the FFA, the PPA, and similarly selective regions, even for the complex stimuli typical of real-world viewing. The real question is whether such information is used in the perception of those stimuli, or whether it is epiphenomenal (13). Some relevant evidence is available for the case of the FFA from the study of individuals with focal brain damage. Some of these individuals exhibit deficits only in face perception (i.e., prosopagnosia), with little or no deficit in object recognition, after damage to regions in or near the FFA, suggesting that even if the FFA contains information about nonfaces, this information is not necessary for object perception. Although no published case of acquired prosopagnosia has completely ruled out the existence of any other deficits beyond face perception (14) using the most sensitive tests of object perception (including reaction time measures) (15), some cases come close (16–19). Note that the rarity of completely clean cases of prosopagnosia without any other deficits is not in itself evidence against the existence of face-specific brain regions, because even if such regions exist, the probability of damaging all and only this region (e.g., in a stroke) is very low.

Because the locus and extent of lesions in humans is not under our control, an importantly complementary method for testing the functional specificity and causal role of cortical regions in perception is transcranial magnetic stimulation (TMS). In TMS, a brief magnetic pulse is delivered to the scalp through a coil held next to the scalp, disrupting neural processing in the cortical region immediately beneath the coil. We can now precisely position the TMS coil to directly target specific cortical regions defined functionally within individual subjects. Although the FFA and PPA are too medial to be reached by TMS, the more lateral face-selective occipital face area (OFA) can be. Using this method, Pitcher et al. (20) showed that TMS to the EBA disrupted perception of bodies (21) but not faces or objects, whereas TMS to the OFA disrupted perception of faces but not objects or bodies. This double dissociation suggests that category-selective regions play a causal role in the perception of their preferred stimulus class but not their nonpreferred stimulus class. Thus, even if the pattern of response across these regions contains some infor-

mation about nonpreferred stimulus categories, the available evidence suggests that such information plays no detectable causal role in perception.

In sum, current evidence suggests that category-selective regions sometimes contain weak but significant information about nonpreferred stimuli, which may be underestimated by fMRI. Nonetheless, results from neuropsychology and TMS are consistent with the hypothesis that any information about nonpreferred stimuli in category-selective regions is epiphenomenal (i.e., not causally involved in perception of those stimuli). It will be important in the future to test this hypothesis further with new data from patients, TMS, and other disruption methods, such as electrical microstimulation in macaque monkeys and humans (22–24).

Why Have Selective Regions in the First Place? Why do some cognitive functions get their own private piece of real estate in the brain, whereas others apparently do not? In thinking about this question, we first need to consider what computational advantages are afforded by functional specialization in the first place. To be detected by fMRI, functional specializations must have two relevant properties: (i) selectivity of the response of neurons to the relevant information (e.g., face selectivity) and (ii) spatial clustering of selective neurons. These phenomena are related but distinct (25) and will be discussed in turn.

Selectivity/sparseness. The advantages of selectivity, or sparseness, in neural coding have been widely noted (26–28). If a given object is coded by the activity of a small subset of the available neurons, then interference is minimized in two important senses. First, it is possible to represent multiple objects simultaneously with minimal ambiguity, because the neural codes for different objects are unlikely to overlap. Thus, we can perceive a face and place simultaneously without the two representations colliding (9). From this perspective, we might expect to find selectivity in neuronal responses when those neurons code for information that we must be able to see despite the simultaneous presence of other visual information. For example, if it is particularly important to be able to detect the presence and identity of another person, without crosstalk from other simultaneously available visual information, it would make sense to allocate special neurons to this job and to make sure the response of those neurons cannot be affected by other stimuli. This possible computational advantage of selectivity is reminiscent of the rationale for the red telephone that was set up between the White House and the Kremlin in the aftermath of the Cuban missile crisis to provide a direct line of communication that was always available and could not be disrupted by transmission of other less important information (4).[‡] Perhaps this is one reason we have neural populations selectively responsive to faces, places, and bodies: to provide private lines of communication about particularly important classes of stimuli that are protected from crosstalk of other irrelevant information.

Second, the use of sparse codes can reduce another form of interference, that caused by learning: we can learn information about one class of objects without altering stored information about another class of objects. With one neural population to represent faces and an overlapping neural population to represent the spatial layout of places, we can learn new faces without disrupting our memories of places and vice versa. From this perspective, we may expect to find relatively sparse codes for classes of information characterized by continual lifelong learning (like faces and places).

[‡]During the crisis, it took 12 h to receive and decode Nikita Khrushchev's 3,000 word initial settlement message—a dangerously long time at such an unstable moment. As noted by Graeme Donald, “at the peak of the crisis the Russian ambassador, Anatoly Dobrynin, was reduced to sending out a man on a bicycle in the middle of the night to collect American replies and then cycle to the nearest Western Union office to relay them to his leader in Moscow. Both sides realized that direct contact was a necessity” (29).

Third, building specialized brain regions, and precise connectivity linking them to other brain regions, could bootstrap development by essentially hardwiring constraints on inductive inference. For example, if information in faces provides the key input required for learning about other people's minds, then perhaps the most efficient way to construct the machinery for thinking about other minds is to hardwire a face area and connect it to another available region, which will then have the constrained input it needs to construct the circuits necessary for social cognition. Evidence against this particular hypothesis comes from the recent finding that congenitally blind individuals show the same location and pattern of activation as sighted subjects when thinking about other people's thoughts, although input from the FFA is likely very different or nonexistent in these people (30). Nevertheless, the general idea that specialized brain regions and their connections may serve as constraints on development is worth considering in other cases.

A fourth possible advantage of relatively sparse codes is metabolic rather than computational: less energy is required if fewer neurons are firing. From this perspective, the greatest lifelong energy savings would come about if sparse codes were available for classes of stimuli that occur most frequently (26). Thus, even from this metabolic perspective, it makes sense to use relatively sparse codes for faces, places, bodies, and words, because they are among the most frequently encountered visual stimuli.

In sum, sparse codes, in which information is coded by a relatively small percentage of the available neurons, each with relatively high selectivity, have certain advantages. At the same time, sparse codes have well-known disadvantages, such as greater susceptibility to damage (because of the smaller number of neurons involved in any given representation) and a smaller number of possible patterns that can be held (one at a time) by a fixed number of neurons. The speculation here is that these disadvantages are outweighed by the particular advantages in the coding of biologically important stimuli like faces, places, and bodies: (i) reduction of interference or crosstalk when multiple stimuli must be represented simultaneously, (ii) the ability to learn new information about one stimulus class without disrupting stored information about another class, and (iii) the potential energy efficiency of coding the most frequently encountered stimuli through the activity of the smallest number of neurons.

Spatial clustering. The second property implied by functionally selective regions detected by fMRI, after selectivity of neurons, is spatial clustering of those neurons. Spatial clustering of functional properties is a familiar phenomenon in the brain, found not only in retinotopic, somatotopic, tonotopic, and other topic maps that follow the organization of the receptor surface, but also in the organization of functional information that is computed de novo, like orientation columns in primary visual cortex and chromotopic maps in posterior inferotemporal cortex (31, 32). Spatial organization is such a pervasive and familiar property of the cortex that we can easily forget to ask ourselves why it occurs. This mystery has been articulated most clearly as follows: "[i]magine taking a cortical area containing a map and scrambling neurons in that area, while preserving all the connections between neurons. Because the circuit remains unchanged, the functional properties of the neurons remain intact. Then the scrambled region without a map is functionally identical to the original one with the map." (33) Given that the identical circuit can be constructed in a spatially clustered or spatially scrambled version, why does spatial clustering occur?

This question is sharpened by two facts: (i) the strong spatial clustering seen in some systems, such as orientation-selective cells in cat visual cortex, is not found in other very similar systems, such as orientation-selective cells in rodent visual cortex (34), and (ii) in the rodent olfactory processing pathway, the precise spatial clustering (and odorant specificity) constructed in the olfactory

bulb, is thrown away in the next stage of processing, the piriform cortex (35).

Chlovskii and Koulakov (33) argue that the need to minimize wiring length (for developmental, metabolic, and conduction delay reasons) must be a fundamental constraint in the nervous system that produces spatial clustering of neurons that are densely connected to each other. To the extent that this wiring-length minimization principle is an important determinant of cortical organization, it suggests that we may find functional specialization in focal cortical regions for functions that are implemented in circuits for which the neurons are densely connected to each other. A testable prediction of this idea is that neurons within face-selective patches of monkey cortex must be richly interconnected, either directly or through webs of inhibitory interneurons found in those same regions. A further prediction of the axon-length minimization principle is that to the extent that readout of a neural code (by the next stage of processing) requires convergence of multiple inputs on a particular neuron, it may be easier to read out a population code represented in a focal region of cortex where those inputs can all conveniently converge on a common output neuron. In a different vein, the functional significance of spatial clustering in the cortex may derive from the requirement to selectively modulate a given functional circuit by way of nonsynaptic diffusible messenger molecules that can spread a few millimeters through the cortex (33).

Functionally specific cortical regions for computationally different problems? None of the computational or biological advantages of neural selectivity and spatial clustering just discussed implies any fundamental difference in the way one object category, coded in a functionally specific cortical region, is represented compared with another. Rather, clustered selectivity is seen as a generally advantageous way of coding any perceptual information. Thus, the observed clustered selectivity for faces, places, bodies, and words need not imply qualitative differences in the computations and representations entailed in the perception of one of these categories versus another. Instead, we might have functional selectivity of the relevant neuronal responses for each category, without any fundamental differences in the kinds of computations conducted for each, just as we see in retinotopic cortex, where completely nonoverlapping pools of neurons code for visual information in one visual field location versus another, but fundamentally similar computations are conducted by each. Continuing this line of thinking, the face, place, and body areas could be seen as subregions of a very large cortical map of all of object space that subsumes all of these regions (5, 36), just as subregions of V1 are parts of a broader representation of retinotopic space. On this view, the deepest question would be not how each of these regions differ from each other computationally, but rather what the dimensions are of that broader space, and hence, why each region lands where it does in that space (36).

Although the mere existence of functional specificity in the cortex for faces, places, bodies, and words does not in itself imply qualitative differences in the processes conducted on each of these different stimulus classes, neither do such arguments for specificity (or their refutations) preclude deeper differences in the computations performed on these stimulus classes. Indeed, especially for the case of faces and places, both theoretical considerations and extensive empirical evidence suggest that different kinds of representations are extracted from these stimulus classes and different uses are made of the resulting information (37–39). A crucial question for the enterprise of using functional specificity of the brain to infer fundamental components of the mind will, therefore, be: which cortical selectivities reflect fundamentally different underlying cognitive processes and which simply reflect convenient compartmentalization of similar processes?

1. Finger S (2001) *Origins of Neuroscience: A History of Explorations into Brain Function* (Oxford University Press, New York).
2. Zola-Morgan S (1995) Localization of brain function: The legacy of Franz Joseph Gall (1758–1828). *Annu Rev Neurosci* 18:359–383.
3. Van Wyhe J (2002) The authority of human nature: The Schädellehre of Franz Joseph Gall. *Br J Hist Sci* 35:17–42.
4. Luzzatti C, Whitaker H (2001) Jean-Baptiste Bouillaud, Claude-François Lallemand, and the role of the frontal lobe: location and mislocation of language in the early 19th century. *Arch Neurol* 58:1157–1162.
5. Kanwisher NG, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
6. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
7. Spiridon M, Kanwisher N (2002) How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35:1157–1165.
8. O'Toole AJ, Jang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. *J Cogn Neurosci* 17:580–590.
9. Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol* 17:2067–2072.
10. Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci USA* 105:4447–4452.
11. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.
12. Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97.
13. Williams MA, Berberovic N, Mattingley JB (2007) Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopagnosia. *Curr Biol* 17:1259–1264.
14. Garrido L, Duchaine B, Nakayama K (2008) Face detection in normal and prosopagnosic individuals. *J Neuropsychol* 2:119–140.
15. Gauthier I, Behrmann M, Tarr MJ (1999) Can face recognition really be dissociated from object recognition? *J Cogn Neurosci* 11:349–370.
16. McNeil JE, Warrington EK (1993) Prosopagnosia: A face-specific disorder. *Q J Exp Psychol A* 46:1–10.
17. Sergent J, Signoret JL (1992) Implicit access to knowledge derived from unrecognized faces in prosopagnosia. *Cereb Cortex* 2:389–400.
18. Wada Y, Yamamoto T (2001) Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *J Neurol Neurosurg Psychiatry* 71:254–257.
19. Riddoch MJ, Johnston RA, Bracewell RM, Boutsen L, Humphreys GW (2008) Are faces special? A case of pure prosopagnosia. *Cogn Neuropsychol* 25:3–26.
20. Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B (2009) Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr Biol* 19:319–324.
21. Urgesi C, Berlucchi G, Aglioti SM (2004) Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Curr Biol* 14:2130–2134.
22. Puce A, Allison T, McCarthy G (1999) Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb Cortex* 9:445–458.
23. Mundel T, et al. (2003) Transient inability to distinguish between faces: Electrophysiologic studies. *J Clin Neurophysiol* 20:102–110.
24. Afraz S-R, Kiani R, Esteky H (2006) Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442:692–695.
25. Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433:597–603.
26. Foldiak P, Young MP (1995) *Handbook of Brain Theory and Neural Networks* (MIT Press, Cambridge, MA).
27. Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. *Curr Opin Neurobiol* 14:481–487.
28. Barlow HB (1995) The neuron doctrine in perception. *The Cognitive Neurosciences*, ed Gazzaniga M (MIT Press, Cambridge, MA), pp 415–436.
29. Donald G (2008) *Sticklers, Sideburns and Bikinis: The Military Origins of Everyday Words and Phrases* (Osprey Publishing, New York).
30. Bedny M, Pascual-Leone A, Saxe RR (2009) Growing up blind does not change the neural bases of Theory of Mind. *Proc Natl Acad Sci USA* 106:11312–11317.
31. Conway BR, Moeller S, Tsao DY (2007) Specialized color modules in macaque extrastriate cortex. *Neuron* 56:560–573.
32. Conway BR, Tsao DY (2009) Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proc Natl Acad Sci USA* 106:18034–18039.
33. Chklovskii DB, Koulakov AA (2004) Maps in the brain: What can we learn from them? *Ann Rev Neurosci* 27:369–392.
34. Ohki K, Reid RC (2007) Specificity and randomness in the visual cortex. *Curr Opin Neurobiol* 17:401–407.
35. Stettler DD, Axel R (2009) Representations of odor in the piriform cortex. *Neuron* 63:854–864.
36. Op de Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ (2008) A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cereb Cortex* 18:1676–1694.
37. Mckone EM, Robbins RR (2010) *The Handbook of Face Perception*, eds Calder AC, et al. (Oxford University Press, New York).
38. Hermer L, Spelke E (1996) Modularity and development: The case of spatial reorientation. *Cognition* 61:195–232.
39. Cheng K, Gallistel CR (1984) Testing the geometric power of an animal's spatial representation. *Animal Cognition: Proceedings of the Harry Frank Guggenheim Conference*, eds Roitblat HL, Bever TG, Terrace HS (Erlbaum, Hillsdale, NJ), pp 409–423.

