



Animal models of the human brain: Successes, limitations, and alternatives

Nancy Kanwisher

The last three decades of research in human cognitive neuroscience have given us an initial “parts list” for the human mind in the form of a set of cortical regions with distinct and often very specific functions. But current neuroscientific methods in humans have limited ability to reveal exactly what these regions represent and compute, the causal role of each in behavior, and the interactions among regions that produce real-world cognition. Animal models can help to answer these questions when homologues exist in other species, like the face system in macaques. When homologues do not exist in animals, for example for speech and music perception, and understanding of language or other people’s thoughts, intracranial recordings in humans play a central role, along with a new alternative to animal models: artificial neural networks.

Addresses

Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, United States

Corresponding author: Kanwisher, Nancy (ngk@mit.edu)

Current Opinion in Neurobiology 2025, 90:102969

This review comes from a themed issue on **Systems Neuroscience 2024**

Edited by **Seung-Hee Lee** and **Mehrdad Jazayeri**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online xxx

<https://doi.org/10.1016/j.conb.2024.102969>

0959-4388/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Introduction

How do human thought and experience arise from the activity of neurons? Although a satisfying mechanistic account of these inchoate phenomena still eludes us, we now have a solid foundation upon which to build. Over the last few decades, cognitive neuroscientists have revealed the large-scale functional organization of the human cortex—the key brain structure responsible for perception and cognition—in considerable detail. Each of us has the same set of functionally distinct brain regions, some generically engaged across content domains and others highly specialized for particular mental processes, from perceiving faces or music or navigational affordances to understanding the meaning of a sentence

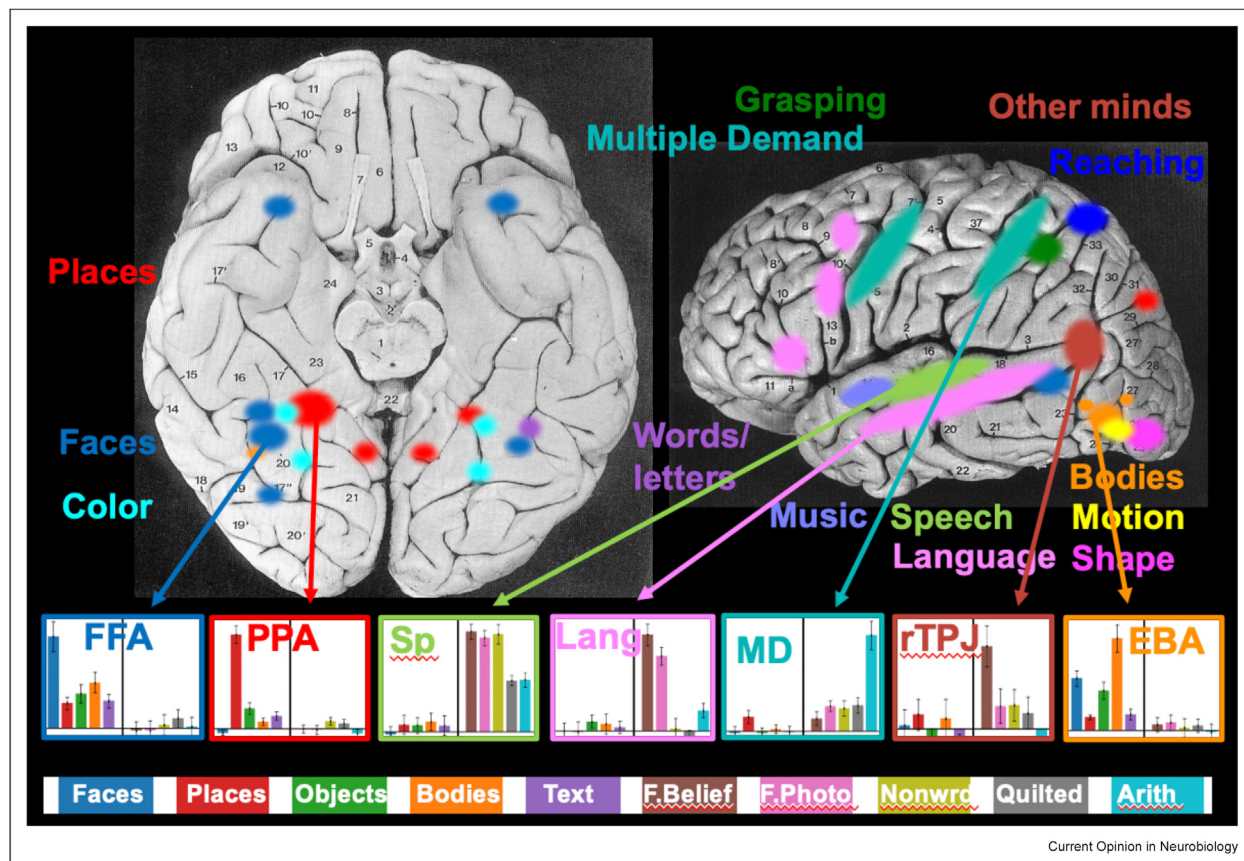
or the nuances of another person’s thoughts ([Figure 1](#)). This “parts list” of the human brain constitutes real progress, but at the same time lays bare a vast space of open questions about the computations, circuit-level mechanisms, and causal role of each of these cortical regions in behavior. To answer those questions, we must overcome the nontrivial methodological shortcomings of human cognitive neuroscience by extrapolating from homologous systems in animals. However, this solution is less helpful for understanding mental functions not shared with animals like language and music and thinking about each other’s thoughts, posing a particular challenge for understanding how neural circuits produce these quintessentially human abilities.

A “parts list” of the human cortex

Unlike subcortical structures, where distinct components are anatomically differentiated in recognizably similar ways across humans and other mammals, subdivisions of the cortex are hard to discern from physical anatomy alone. Primary cortical regions have long been recognized in humans by their distinctive thalamic input or corticospinal tract output, and postmortem histology from Brodmann to modern methods [1,2] indicates some cytoarchitectonic differentiation across the rest of the cortex. But many of these anatomically defined regions do not have sharp borders, and their relation to function could until recently be inferred only imprecisely from the deficits that result in neurological patients with lesions in their approximate location.

This situation changed abruptly with the advent of fMRI in the early 1990s. Retinotopic mapping allowed the identification of V1 and several downstream retinotopic regions, as well as visual motion area MT, all with clear or likely homologues to regions previously studied in detail with electrophysiology in macaques. The functional region of interest (fROI) approach, in which particular fMRI “localizer” contrasts were used to identify fROIs in each participant individually, made possible the study of functionally distinctive cortical regions that varied between subjects in their exact anatomical location [3]. The adoption of the same localizers across labs to identify these regions has made possible a cumulative research program in which the existence and functional and anatomical [4] properties

Figure 1



Functionally specific cortical regions in the human brain. Colored blobs indicate the approximate location of functionally distinctive cortical regions that are present in virtually every subject and that have been widely replicated across labs. Bar charts show the response of several of these regions to passive viewing of five visual categories (faces, places, objects, bodies, and text) presented simultaneously with five auditory task conditions (assessing completions of short verbal descriptions of false beliefs (FB), false photos (FP), nonwords (NW), quilted speech (QS), and arithmetic). Each bar shows the mean response across ten participants in held-out data from those used to define each region (FFA: faces > objects; PPA: places > objects; EBA: bodies > objects; rTPJ: FB > FP; Language Regions: EP & PP > NW; Speech regions: NW > QS; MD: arithmetic > NW). Because visual and auditory conditions were presented simultaneously (and orthogonalized), bar charts for visual regions have the mean response across all visual conditions subtracted from each auditory condition, and bar charts for other regions have the mean response across auditory conditions subtracted from each visual condition.

of each of these regions have been widely replicated and characterized in detail.

Using the fROI method, a number of category-selective regions were identified in the ventral visual pathway that had not been previously described in macaques, including the fusiform face area, parahippocampal place area, extrastriate body area, and visual word form area, responding selectively to faces, places, bodies, and text, respectively [5]. Although multiple voxel pattern analysis (MVPA) enables nonpreferred information to be extracted from fMRI responses in each of these regions, and weaker selectivities for real-world size, animacy, and eccentricity overlap with many of these regions [6], the available causal evidence from neurological patients, TMS, and intracranial stimulation supports a selective role of each region in the perception of stimuli from its preferred category (see Box 1).

Similar methods have identified cortical regions with selective roles in distinctively human functions including the perception of speech sounds and music [7,8]. Further, resolving a centuries-long debate, it is now clear that brain regions engaged in language processing are highly selective for language per se, revealing little or no response when people perform mental arithmetic, solve logic problems, hold information in working memory, listen to music, or exert cognitive control [9]. Although language-selective regions in the frontal lobe reside near functionally less specific “multiple demand” regions, the individual-subject fROI method has shown these regions to be nonoverlapping at the individual-subject level [9]. Thus, language and thought are distinct in the brain [10]. In perhaps the most astonishing functionally specific region of the cortex, Saxe and colleagues have shown that a region of the temporo-parietal junction is

Box 1

The most significant critique of the functional specificity of category-selective regions is the discovery from Haxby [79] and others that it is possible to decode information about nonpreferred categories from the response across voxels in these regions (e.g. discriminating the response to shoes versus cars in the FFA). However, this nonpreferred information is weak, and Haxby himself noted that “preferred regions for faces ... are not well suited to object classifications that do not involve faces ...” [80]. Further, the mere existence of information about nonfaces in the FFA is unsurprising, as the projection of different nonface stimuli into face space can be expected to yield at least partly different representations. Indeed, ANNs trained only on face discrimination, which can be thought of as the Platonic ideal of a face-specific system, can decode non faces above chance [60,81]. More fundamentally, the fact that scientists can read out weak nonpreferred information from category-specific regions does not mean the rest of the brain is reading out this information. The key question is whether this information plays a causal role in behavior. Most available evidence from interventions such as lesions [82], pharmacological [36], or electrical disruption [42,43] restricted to face-specific regions supports a specific causal role in face perception (but see Ref. [37]).

selectively engaged when people think about what another person is thinking [11], not when they think about a person’s physical characteristics, social practices, or even their bodily sensations like thirst or hunger or pain.

Although some have argued against the idea that many cortical regions are specifically engaged in particular mental functions [12,13] their critiques have not engaged with and cannot account for the evidence for the functional selectivity of these regions from fMRI, intracranial recordings, and specific deficits after focal brain damage, TMS, and direct electrical stimulation (see also [14]). For a more substantive critique of functional specificity and its counterarguments, see Box 1.

The power of animal models: the face system

Thus, fMRI has proven a powerful tool for discovering a set of functionally distinctive cortical regions we all share. However, a rich understanding of these regions requires not just a characterization of their selectivity, but an of understanding the information represented and computations conducted in each region, and how these representations and computations are implemented in actual circuits of neurons. fMRI can take some initial steps to address these questions using methods like MVPA to characterize the information content in each region, and multivariate analyses of EEG and MEG to detect changes in representational content over time. But methods in humans provide little ability to precisely characterize neural representations, watch them evolve over time to reveal the process of computation, discover the actual structural connectivity of

remote cortical regions, chart information flow between them, or test the causal role of any of these phenomena in behavior. All of these questions can be answered in animal models. Face perception provides a parade case of this synergy between human and animal studies.

The discovery of the fusiform face area (FFA) in humans was foreshadowed by prior findings of “face cells” in monkey IT cortex [15], as well as reports of human patients with selective deficits in face perception after damage to the right inferior temporal lobe. Nonetheless, the ability to identify the FFA in essentially each individual using the fROI method described above made possible a systematic research program of characterizing the response profile of this region in detail. Early research showed that the FFA could not be easily accounted for in terms of a selectivity for low-level visual features such as curvature, but instead reflected actual selectivity for faces themselves. This finding supports the long-standing idea of domain specificity in cognitive science [16], according to which many mental functions are computed in cognitive/neural machinery highly specialized for processing that domain of information. Further studies showed that activity in the FFA was correlated with perceptual awareness of faces even when the stimulus did not change (in binocular rivalry) [17], activated by just imagining faces with eyes closed [18], and sensitive to face identity in upright but not inverted faces (echoing long-standing behavioral findings [19]). More recent work has shown that the FFA is already present in five-month-old infants [20,21].

In addition to the FFA, other face-selective regions have been found in humans, including the occipital face area (OFA), which encodes more image-based features of faces, a face-selective region in the pSTS which is particularly sensitive to dynamic information in faces and also responds to human voices, and face-selective regions have been reported in more anterior temporal regions and in the frontal lobes [22]. A recent paper found higher responses to familiar than unfamiliar faces in the temporal pole and perirhinal cortex [23].

Most excitingly, Tsao, Freiwald, and their collaborators identified multiple face-selective patches in macaques, and then went far beyond anything possible in humans. They recorded from individual neurons in identified patches and found that nearly all cells in the middle face patch were selectively responsive to faces [24], they characterized the progression of representations from view-specific to mirror-symmetric viewpoint tuning to fully viewpoint invariant tuning to individual identity across face patches [25], and they showed the strict connectivity [26] between face patches but not between those patches and the intervening cortical tissue. Further, they precisely described the selectivities of individual neurons in face patches [27–29], leading to a characterization of the neural code for face identity [30],

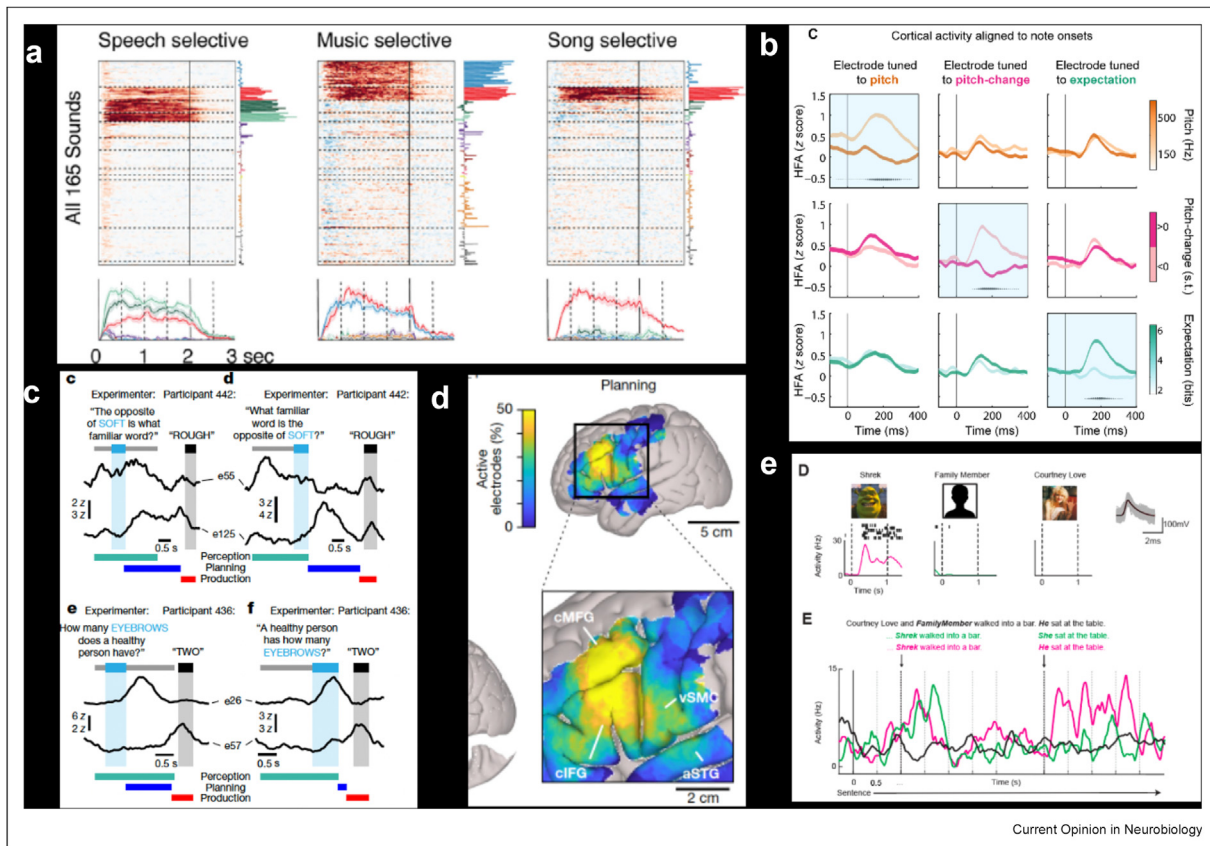
and they identified the particular face patches [31,32] and coding schemes [33] that distinguish between familiar and unfamiliar faces.

Although the face system in general is likely homologous between macaques and humans, the precise correspondence between face patches is not clear [34], and a strict one-to-one mapping between regions across species may not exist. These questions may be answered in the future by a large ongoing multisite project that is attempting to combine functional MRI imaging in vivo with postmortem single-cell RNA sequencing in humans and primates to discover whether functionally specific regions such as the FFA have distinct cell type signatures, and if so if these are shared across species. But even if one-to-one homologies are not ultimately found between species, the general findings from the face system in macaques are likely to apply to humans, given the similarities between species in visual

processing behavior [35] and in the location of the face system across species relative to other functional landmarks. The work in primates is thus a great gift to those of us trying to understand the neural basis of face perception in humans, answering many questions that are difficult or impossible to answer in humans.

That said, some advantages of human studies remain. One example concerns causal tests of the role of face patches in behavior. Studies in macaques have demonstrated alterations of face perception when face patches are disrupted electrically, pharmacologically, or optogenetically [36–39]. But each of these experiments shows a reduction in accuracy on a single highly trained task, a low-dimensional reflection of what is surely a very high dimensional change in the monkey’s percept. A clever new method called “perceptography” [40] can depict the contents of a monkey’s percept induced by a local perturbation of the IT cortex, but requires

Figure 2



Intracranial recordings reveal neural mechanisms of uniquely human speech, language, and music processing. a. Three electrodes in human auditory cortex are selectively responsive to speech, music, and song, respectively [8]. Top: gamma power over time in response to each of 165 natural sounds containing foreign speech (light green), English speech (dark green), instrumental music (blue), and vocal music (red). Bottom: gamma power over time averaged within each sound category. b. Three electrodes in human auditory cortex are selectively responsive to pitch, pitch change, and expectation in music [50]. c. Responses of individual electrodes during conversation, revealing selective responses during speech perception, speech planning, and speech production [54]; Timing of onset of speech planning is manipulated by placing the critical word (“soft” or “eyebrows” in these examples) in different positions in the question the subject is asked. d. Cortical location of speech planning responses, distinct from cortical loci for speech perception and production [54]. e. Response of an electrode in human hippocampus that responds selectively to the image or spoken name of Shrek, and also to a pronoun referring to Shrek, revealing neural correlates of referent retrieval in online sentence comprehension [83].

thousands of trials per site. In contrast, on the rare occasions when neurosurgeons electrically stimulate face-specific sites in humans for clinical reasons, these patients have done something no monkey can, by using language to share with us rich insights into their resulting percepts (see also [41]). One patient reported during stimulation of the FFA that “your face just metamorphosed ... that was a trip!” [42] and another reported “your face kind of changes into an anime character” and for stimulation of the OFA he said, “I start seeing eyes over here, but if I look over here, I see eyes over here” [43].

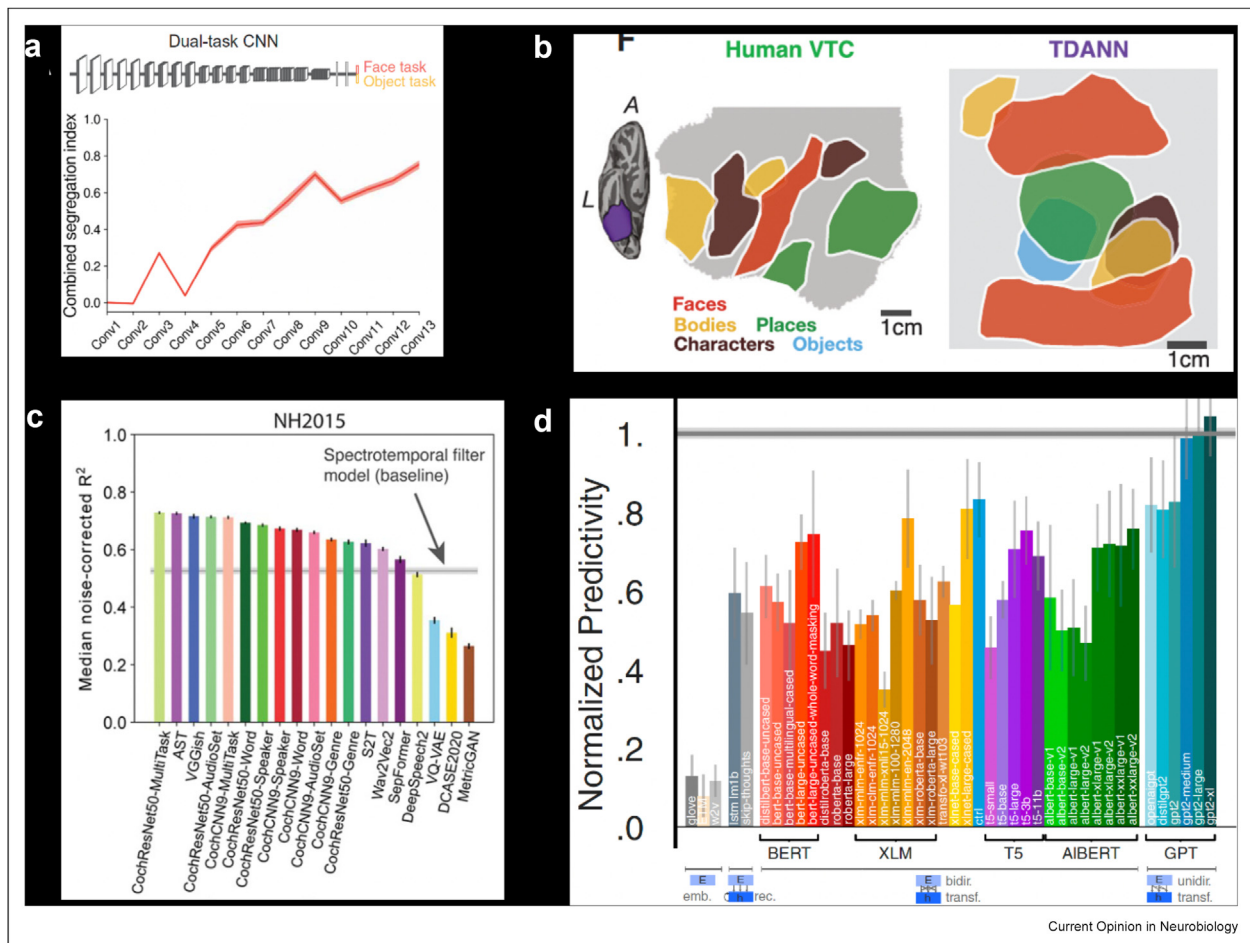
What do we do when no animal models exist?

The progress in understanding the brain basis of face perception illustrates the power of combining complementary studies of a particular cortical system in

humans and monkeys. Similar efforts are ongoing to understand brain systems that process other domains that are likely largely shared between humans and macaques such as body perception [44], scene perception [45], the perception of social interactions [46,47], and intuitive physical reasoning [47,48]. But how are we to understand domains of perception and cognition that are not shared with other animals, such as the perception of speech or music sounds, or understanding the meaning of a sentence or the contents of another person’s thoughts? Nonhuman primates lack these abilities, posing a particular challenge for studying them at the level of neural populations and circuits.

For brain regions that implement uniquely human functions, a particularly important source of data for fine-grained neural understanding comes from rare and precious intracranial recordings from neurosurgery

Figure 3



Artificial neural networks as computational models of human brain functions. a. CNNs trained on both object classification and face identification spontaneously segregate themselves into two distinct systems, to an increasing degree in later layers of the network, mirroring the functional organization of the brain [81]. b. A topographic deep artificial neural network predicts category-selective regions much like those observed in the brain [65]. c. Audio-trained neural networks, predict voxelwise responses to natural sounds measured with fMRI [70]. d. Large Language models, especially GPT models, predict voxelwise responses to sentences in the human brain [72].

patients with implanted electrodes (see Figure 2). These recordings have already revealed neural populations specific for speech and music [8], and have begun to unpack the neural codes underlying each [49–51]. Other intracranial recordings in the language cortex have revealed glimpses into the computations entailed in sentence understanding [52,53] and production [54]. These recordings only rarely provide a high density of electrodes in a cortical region of interest [8,49], as clinical needs determine their use, and most neurosurgeons now use depth electrodes that sample the brain widely but sparsely. Although several new high-density recording arrays have been developed recently for humans [55–57], these have only been used for brief periods during surgery. It would be game-changing for human cognitive neuroscience if clinically useful high-density recording devices were developed that could be safely used for longer periods.

Complementary approaches using ANNs

In an exciting discovery over the last decade, artificial neural networks (ANNs) trained on natural stimuli have been found to perform well on humanlike tasks, meaning that they can serve as computationally explicit hypotheses of how these problems might be solved in the brain. Indeed, ANNs recapitulate many behavioral and neural phenomena observed in primates, from visual and auditory psychophysical effects [58–61] to neural responses [62,63], including the mirror-symmetric viewpoint tuning of face cells found previously in a mid-level monkey face patch [64], and even cortical topography in the primate ventral visual pathway [65,66]. The spontaneous emergence of these phenomena in networks not designed to match primate data (Figure 3, top), but merely optimized for task performance, points to deep computational reasons why these phenomena are observed in primates [67,68]. Further, ANN models are also capturing many aspects of distinctively human phenomena for which we do not have good animal models (Figure 3, bottom), including the perception of pitch [69], neural responses to speech sounds and music [70,71], and behavioral and neural responses during language comprehension [52,72–76].

The ability of ANNs to perform well on human tasks and to capture many aspects of human behavior and neural responses means that they can serve as useful models of the computations performed by the brain. Although obviously different in their physical implementation, ANNs have the advantages that they give complete control over the data the models are trained on (i.e. for controlled rearing studies), they provide full information about the response of each unit in the network to any input, and they enable precise causal interventions on the networks in any way we wish. ANNs thus powerfully complement animal models of

human functions where interspecies homologies exist, and fill in for the missing animal models of uniquely human cognition [70–72,77,78].

Conclusions

In sum, human cognitive neuroscience has made substantial progress sketching the large-scale functional organization of the cortex in humans. Where homologous cortical regions exist in animals, they can provide finer-grained characterizations of neural codes and computations, and how they change over time. Where animal models don't exist, we will have to be smarter, analogizing from systems where we do have good animal models, making maximal use of rare intracranial data from humans, and using artificial neural networks as substitutes for animal models. But what we really need are better methods in humans that can i) reveal the responses of large numbers of neurons in multiple cortical regions with precise time information, ii) discover actual structural connectivity between cortical regions, and iii) watch information flow from one brain area to the next.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I think Rebecca Saxe, Greta Tuckute, and Doris Tsao for helpful comments, Sam Hutchinson for analyzing the data to produce the bar charts in Figure 1, and Ammar Marvi for managing the references. This work was funded by NIH grants EY033843 and UM1MH130981.

Data availability

This is a review article which does not present its own new data.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Amunts K, Zilles K: **Architectonic mapping of the human brain beyond Brodmann.** *Neuron* 2015, **88**:1086–1107.
2. Petersen SE, Seitzman BA, Nelson SM, Wig GS, Gordon EM: **Principles of cortical areas and their implications for neuroimaging.** *Neuron* 2024, <https://doi.org/10.1016/j.neuron.2024.05.008>.
3. Saxe R, Brett M, Kanwisher N: **Divide and conquer: A defense of functional localizers.** *Neuroimage* 2006, **30**:1088–1096.
4. Rosenke M, Weiner KS, Barnett MA, Zilles K, Amunts K, Goebel R, Grill-Spector K: **A cross-validated cytoarchitectonic atlas of the human ventral visual stream.** *Neuroimage* 2018, **170**:257–270.
5. Kanwisher N: **Functional specificity in the human brain: a window into the functional architecture of the mind.** *Proc Natl Acad Sci U S A* 2010, **107**.

6. Konkle T, Oliva A: **A Real-World Size Organization of Object Responses in Occipitotemporal Cortex.** *Neuron* 2012, **74**: 1114–1124.
7. Norman-Haignere S, Kanwisher NG, McDermott JH: **Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition.** *Neuron* 2015, **88**: 1281–1296.
8. Norman-Haignere SV, Feather J, Boebinger D, Brunner P, Ritaccio A, McDermott JH, Schalk G, Kanwisher N: **A neural population selective for song in human auditory cortex.** *Curr Biol* 2022, **32**:1454–1455.
9. Fedorenko E, Ivanova AA, Regev TI: **The language network as a natural kind within the broader landscape of the human brain.** *Nat Rev Neurosci* 2024, **25**:289–312.
10. Fedorenko E, Varley R: **Language and thought are not the same thing: evidence from neuroimaging and neurological patients.** *Ann N Y Acad Sci* 2016, **1369**:132–153.
11. Saxe R: **The right temporo-parietal junction: a specific brain region for thinking about thoughts.** In *Handbook of theory of mind*; 2010:1–35.
12. Pessoa L: **The entangled brain.** *J Cognit Neurosci* 2023, **35**: 349–360.
13. Westlin C, Theriault JE, Katsumi Y, Nieto-Castanon A, Kucyi A, Ruf SF, Brown SM, Pavel M, Erdogmus D, Brooks DH, et al.: **Improving the study of brain-behavior relationships by revisiting basic assumptions.** *Trends Cognit Sci* 2023, **27**:246–257.
14. Krakauer JW: **Modular Brain, Entangled Argument.** *J Cognit Neurosci* 2023, **35**:388–390.
The author critiques the view that because cognition is complex must take place in distributed networks whose nodes switch functions dynamically between tasks.
15. Desimone R: **Face-Selective Cells in the Temporal Cortex of Monkeys.** *J Cognit Neurosci* 1991, **3**:1–8.
16. Hirschfeld LA, Gelman SA: **Toward a topography of mind: an introduction to domain specificity.** In *Mapping the mind.* Cambridge University Press; 1994:3–36.
17. Tong F, Nakayama K, Vaughan JT, Kanwisher N: **Binocular Rivalry and Visual Awareness in Human Extrastriate Cortex.** *Neuron* 1998, **21**:753–759.
18. O'Craven KM, Kanwisher N: **Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions.** *J Cognit Neurosci* 2000, **12**:1013–1023.
19. Yin RK: **Looking at upside-down faces.** *J Exp Psychol* 1969, **81**: 141–145.
20. Kosakowski HL, Cohen MA, Herrera L, Nichoson I, Kanwisher N, Saxe R: **Cortical Face-Selective Responses Emerge Early in Human Infancy.** *eNeuro* 2024, <https://doi.org/10.1523/ENEURO.0117-24.2024>.
21. Kosakowski HL, Cohen MA, Takahashi A, Keil B, Kanwisher N, Saxe R: **Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants.** *Curr Biol* 2022, **32**: 265–274.e5.
22. Nikel L, Sliwinska MW, Kucuk E, Ungerleider LG, Pitcher D: **Measuring the response to visually presented faces in the human lateral prefrontal cortex.** *Cereb Cortex Commun* 2022, **3**.
23. Deen B, Husain G, Freiwald WA: **A familiar face and person processing area in the human temporal pole.** *Proc Natl Acad Sci USA* 2024, **121**.
Using fMRI this paper finds higher responses to familiar than unfamiliar faces in human temporal pole (TP) and perirhinal cortex (PR), and further finds that TP but not PR is further engaged in other high-level social cognitive tasks that involve thinking about people including about their thoughts.
24. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS: **A Cortical Region Consisting Entirely of Face-Selective Cells.** *Science* (1979) 2006, **311**:670–674.
25. Freiwald WA, Tsao DY: **Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System.** *Science* (1979) 2010, **330**:845–851.
26. Moeller S, Freiwald WA, Tsao DY: **Patches with Links: A Unified System for Processing Faces in the Macaque Temporal Lobe.** *Science* (1979) 2008, **320**:1355–1359.
27. Freiwald WA, Tsao DY, Livingstone MS: **A face feature space in the macaque temporal lobe.** *Nat Neurosci* 2009, **12**:1187–1196.
28. Ohayon S, Freiwald WA, Tsao DY: **What Makes a Cell Face Selective? The Importance of Contrast.** *Neuron* 2012, **74**: 567–581.
29. Yang Z, Freiwald WA: **Encoding of dynamic facial information in the middle dorsal face area.** *Proc Natl Acad Sci USA* 2023, **120**.
The middle dorsal face area (MD) in macaques is shown to encode not only face shape but also complex naturalistic face motion.
30. Chang L, Tsao DY: **The Code for Facial Identity in the Primate Brain.** *Cell* 2017, **169**:1013–1028. e14.
31. Landi SM, Freiwald WA: **Two areas for familiar face recognition in the primate brain.** *Science* (1979) 2017, **357**:591–595.
32. Landi SM, Viswanathan P, Serene S, Freiwald WA: **A fast link between face perception and memory in the temporal pole.** *Science* (1979) 2021, **373**:581–585.
33. She L, Benna MK, Shi Y, Fusi S, Tsao DY: **Temporal multiplexing of perception and memory codes in IT cortex.** *Nature* 2024, **629**:861–868.
34. Freiwald W, Duchaine B, Yovel G: **Face Processing Systems: From Neurons to Real-World Social Perception.** *Annu Rev Neurosci* 2016, **39**:325–346.
35. Rajalingham R, Schmidt K, DiCarlo JJ: **Comparison of Object Recognition Behavior in Human and Monkey.** *J Neurosci* 2015, **35**:12127–12136.
36. Sadagopan S, Zarco W, Freiwald WA: **A causal relationship between face-patch activity and face-detection behavior.** *Elife* 2017, **6**.
37. Moeller S, Crago T, Chang L, Tsao DY: **The effect of face patch microstimulation on perception of faces and objects.** *Nat Neurosci* 2017, **20**:743–752.
38. Afraz S-R, Kiani R, Esteky H: **Microstimulation of inferotemporal cortex influences face categorization.** *Nature* 2006, **442**: 692–695.
39. Afraz A, Boyden ES, DiCarlo JJ: **Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination.** *Proc Natl Acad Sci USA* 2015, **112**:6730–6735.
40. Shahbazi E, Ma T, Pernuš M, Scheirer W, Afraz A: **Perceptography unveils the causal contribution of inferior temporal cortex to visual perception.** *Nat Commun* 2024, **15**:3347.
Trained monkeys to detect optical stimulation of sites in IT cortex and then generated images that “fooled” the monkeys into thinking that site was stimulated, producing a “perceptograms” that reveals the perceptual correlate of activation of that site in IT.
41. Herald SB, Almeida J, Duchaine B: **Face distortions in prosopometamorphopsia provide new insights into the organization of face perception.** *Neuropsychologia* 2023, **182**, 108517.
In prosopometamorphopsia (PMO), features appear distorted only for faces, usually without a disruption in face recognition accuracy, addressing many theoretical questions in visual neuroscience.
42. Parvizi J, Jacques C, Foster BL, Withoft N, Rangarajan V, Weiner KS, Grill-Spector K: **Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception.** *J Neurosci* 2012, **32**:14915–14920.
43. Schalk G, Kapeller C, Guger C, Ogawa H, Hiroshima S, Lafer-Sousa R, Saygin ZM, Kamada K, Kanwisher N: **Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain.** *Proc Natl Acad Sci USA* 2017, **114**:12285–12290.

44. Bao P, She L, McGill M, Tsao DY: **A map of object space in primate inferotemporal cortex.** *Nature* 2020, **583**:103–108.
45. Epstein RA, Baker CI: **Scene Perception in the Human Brain.** *Annu Rev Vis Sci* 2019, **5**:373–397.
46. McMahon E, Isik L: **Seeing social interactions.** *Trends Cognit Sci* 2023, **27**:1165–1179.
47. Sliwa J, Freiwald WA: **A dedicated network for social interaction processing in the primate brain.** *Science (1979)* 2017, **356**:745–749.
48. Fischer J, Mikhael JG, Tenenbaum JB, Kanwisher N: **Functional neuroanatomy of intuitive physical inference.** *Proc Natl Acad Sci USA* 2016, **113**.
49. Leonard MK, Gwilliams L, Sellers KK, Chung JE, Xu D, Mischler G, Mesgarani N, Welkenhuysen M, Dutta B, Chang EF: **Large-scale single-neuron speech sound encoding across the depth of human cortex.** *Nature* 2024, **626**:593–602.
Single neurons in high-level auditory cortex, recorded with neuropixels in humans, were sensitive to a variety of speech sounds including features of consonants and vowels, relative pitch, onsets, and amplitude and sequence statistics.
50. Sankaran N, Leonard MK, Theunissen F, Chang EF: **Encoding of melody in the human auditory cortex.** *Sci Adv* 2024, **10**.
Intracranial recordings while people listen to music found sites selective for music that were sensitive to musical expectation and sites sensitive to pitch and pitch-change in both music and speech.
51. Norman-Haignere SV, Long LK, Devinsky O, Doyle W, Irobunda I, Merricks EM, Feldstein NA, McKhann GM, Schevon CA, Flinker A, et al.: **Multiscale temporal integration organizes hierarchical computation in human auditory cortex.** *Nat Human Behav* 2022, **6**:455–469.
Develops a new method to estimate time windows of sensory integration, and applies it to intracranial responses from human auditory cortex, finding some short-integration electrodes (< 200 ms) with spectrotemporal modulation selectivity, and others with long integration windows (> than 200 ms) with strong category selectivity.
52. Desbordes T, Lakretz Y, Chanoine V, Oquab M, Badier J-M, Trébuchon A, Carron R, Bénar C-G, Dehaene S, King J-R: **Dimensionality and Ramping: Signatures of Sentence Integration in the Dynamics of Brains and Deep Language Models.** *J Neurosci* 2023, **43**:5350–5364.
53. Jamali M, Grannan B, Cai J, Khanna AR, Muñoz W, Caprara I, Paulk AC, Cash SS, Fedorenko E, Williams ZM: **Semantic encoding during language comprehension at single-cell resolution.** *Nature* 2024, <https://doi.org/10.1038/s41586-024-07643-2>.
54. Castellucci GA, Kovach CK, Howard MA, Greenlee JDW, Long MA: **A speech planning network for interactive language use.** *Nature* 2022, **602**:117–122.
Intracranial recordings from the human frontal lobe during conversation find neural correlates of speech planning, distinct from both speech perception and speech articulation.
55. Chung JE, Sellers KK, Leonard MK, Gwilliams L, Xu D, Dougherty ME, Kharazia V, Metzger SL, Welkenhuysen M, Dutta B, et al.: **High-density single-unit human cortical recordings using the Neuropixels probe.** *Neuron* 2022, **110**:2409–2421. e3.
56. Paulk AC, Kfir Y, Khanna AR, Mustroph ML, Trautmann EM, Soper DJ, Stavisky SD, Welkenhuysen M, Dutta B, Shenoy KV, et al.: **Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex.** *Nat Neurosci* 2022, **25**:252–263.
57. Lee K, Paulk AC, Ro YG, Cleary DR, Tonsfeldt KJ, Kfir Y, Pezaris JS, Tchoe Y, Lee J, Bourhis AM, et al.: **Flexible, scalable, high channel count stereo-electrode for recording in the human brain.** *Nat Commun* 2024, **15**:218.
58. Rajalingham R, Issa EB, Bashivan P, Kar K, Schmidt K, DiCarlo JJ: **Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks.** *J Neurosci* 2018, **38**:7255–7269.
59. Jacob G, Pramod RT, Katti H, Arun SP: **Qualitative similarities and differences in visual object representations between brains and deep networks.** *Nat Commun* 2021, **12**:1872.
60. Dobs K, Yuan J, Martinez J, Kanwisher N: **Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition.** *Proc Natl Acad Sci USA* 2023, **120**.
61. Franci A, McDermott JH: **Deep neural network models of sound localization reveal how perception is adapted to real-world environments.** *Nat Human Behav* 2022, **6**:111–133.
A deep neural network trained to localize sounds spontaneously exhibited many of the features of human spatial hearing, but only when the model was trained on natural inputs (with reverb), not when trained on unnatural inputs without reverb, showing how artificial neural networks can reveal the task constraints that shape human perception.
62. Yamins DLK, DiCarlo JJ: **Using goal-driven deep learning models to understand sensory cortex.** *Nat Neurosci* 2016, **19**:356–365.
63. Ratan Murty NA, Bashivan P, Abate A, DiCarlo JJ, Kanwisher N: **Computational models of category-selective brain regions enable high-throughput tests of selectivity.** *Nat Commun* 2021, **12**:5540.
64. Farzmaidi A, Zarco W, Freiwald WA, Kriegeskorte N, Golan T: **Emergence of brain-like mirror-symmetric viewpoint tuning in convolutional neural networks.** *Elife* 2024, **13**.
Echoing the progression of view tuning across the cortical hierarchy of macaque face patches, convolutional neural networks trained to discriminate among bilaterally symmetric objects show a shift from view specificity to mirror symmetry to full view invariance across layers.
65. Margalit E, Lee H, Finzi D, DiCarlo JJ, Grill-Spector K, Yamins DLK: **A unifying framework for functional organization in early and higher ventral visual cortex.** *Neuron* 2024, <https://doi.org/10.1016/j.neuron.2024.04.018>.
Topographic deep artificial neural networks trained unsupervised on natural images predict category-specific cortical regions.
66. Blauch NM, Behrmann M, Plaut DC: **A connectivity-constrained computational account of topographic organization in primate high-level visual cortex.** *Proc Natl Acad Sci USA* 2022, **119**.
67. Kanwisher N, Khosla M, Dobs K: **Using artificial neural networks to ask ‘why’ questions of minds and brains.** *Trends Neurosci* 2023, **46**:240–254.
68. Khosla M, Williams AH, McDermott J, Kanwisher N: **Privileged representational axes in biological and artificial neural networks.** *bioRxiv* 2024, <https://doi.org/10.1101/2024.06.20.599957>.
69. Saddler MR, Gonzalez R, McDermott JH: **Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception.** *Nat Commun* 2021, **12**:7278.
70. Tuckute G, Feather J, Boebinger D, McDermott JH: **Many but not all deep neural network audio models capture brain responses and exhibit correspondence between model stages and brain regions.** *PLoS Biol* 2023, **21**, e3002366.
Audio neural network models, especially those trained on multiple tasks, outpredict spectrotemporal filter-bank models of fMRI responses in auditory cortex, showing the promise of deep neural network models of audition.
71. Li Y, Anumanchipalli GK, Mohamed A, Chen P, Carney LH, Lu J, Wu J, Chang EF: **Dissecting neural computations in the human auditory pathway using deep neural networks for speech.** *Nat Neurosci* 2023, **26**:2213–2225.
Representations in deep neural network models trained on speech input show a brain-like hierarchy, and networks trained on either English or Mandarin predicted cortical responses in native speakers of each language.
72. Schrimpf M, Blank IA, Tuckute G, Kauf C, Hosseini EA, Kanwisher N, Tenenbaum JB, Fedorenko E: **The neural architecture of language: integrative modeling converges on predictive processing.** *Proc Natl Acad Sci USA* 2021, **118**.
73. Goldstein A, Grinstein-Dabush A, Schain M, Wang H, Hong Z, Aubrey B, Schain M, Nastase SA, Zada Z, Ham E, et al.: **Alignment of brain embeddings and artificial contextual embeddings in natural language points to common geometric patterns.** *Nat Commun* 2024, **15**:2768.
74. Caucheteux C, King J-R: **Brains and algorithms partially converge in natural language processing.** *Commun Biol* 2022, **5**:134.

75. Antonello R, Huth A: **Predictive Coding or Just Feature Discovery? An Alternative Account of Why Language Models Fit Brain Data.** *Neurobiology of Language* 2023, https://doi.org/10.1162/nol_a_00087.
76. Kumar S, Sumers TR, Yamakoshi T, Goldstein A, Hasson U, Norman KA, Griffiths TL, Hawkins RD, Nastase SA: **Shared functional specialization in transformer-based language models and the human brain.** *Nat Commun* 2024, **15**:5523.
77. Kauf C, Tuckute G, Levy R, Andreas J, Fedorenko E: **Lexical-Semantic Content, Not Syntactic Structure, Is the Main Contributor to ANN-Brain Similarity of fMRI Responses in the Language Network.** *Neurobiology of Language* 2024, **5**:7–42.
- This paper asks what aspects of language input explain the ability of LLMs to predict neural responses by comparing neural predictions from intact language input to input scrambled in word order, input with words missing, or input with varying semantic similarity to the original sentence, and finds that LLM fit to the brain is driven more by semantic than structural properties of the input.
78. Tuckute G, Sathe A, Srikant S, Taliaferro M, Wang M, Schrimpf M, Kay K, Fedorenko E: **Driving and suppressing the human language network using large language models.** *Nat Human Behav* 2024, **8**:544–561.
- Used a GPT-based encoding model of language cortex to identify sentences predicted to drive or suppress the response, and empirically validated these predictions. Found that surprisal and well-formedness of linguistic input are key determinants of response strength in the language network.
79. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P: **Distributed and overlapping representations of faces and objects in ventral temporal cortex.** *Science (1979)* 2001, **293**: 2425–2430.
80. O'Toole AJ, Jiang F, Abdi H, Haxby JV: **Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex.** *J Cognit Neurosci* 2005, **17**:580–590.
81. Dobs K, Martinez J, Kell AJE, Kanwisher N: **Brain-like functional specialization emerges spontaneously in deep neural networks.** *Sci Adv* 2022, **8**.
82. Wada Y, Yamamoto T: **Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region.** *J Neurol Neurosurg Psychiatry* 2001, **71**: 254–257.
83. Dijksterhuis DE, Self MW, Possel JK, Peters JC, van Straaten ECW, Idema S, Baaijen JC, van der Salm SMA, Aarnoutse EJ, van Klink NCE, *et al.*: **Pronouns reactivate conceptual representations in human hippocampal neurons.** *bioRxiv* 2024, <https://doi.org/10.1101/2024.06.23.600044>.