OPINION

A new look at domain specificity: insights from social neuroscience

Robert P. Spunt and Ralph Adolphs

Abstract | The concept of domain specificity — which suggests that some aspects of neural processing are specialized for particular types of stimuli — has been invoked to explain a range of cognitive phenomena, including language, face perception and theory of mind, and has been a hallmark of theories of cognitive architecture. More recent usage of this concept draws on neuroscientific data and, in particular, on work in social neuroscience. A critical examination of the part that the concept of domain specificity has played in theories of human brain function leads us to suggest a new view according to which domain specificity pertains to centrally generated constraints on information processing that can be both dynamic and context sensitive.

Gross dissection of a brain immediately suggests that different parts of the brain do different things: brains have architecture. Indeed, we now know that, even though computation is spatially distributed within the brain in complicated ways, different neural structures are concerned to some degree with processing different kinds of information. The brain is certainly not equipotential. However, there remain a number of interesting and difficult questions about the degree of such apparent specialization, how it might come about and what it accomplishes. These issues have been the focus of numerous theories of domain specificity, which range from abstract cognitive hypotheses to neurophysiological and neuroanatomical accounts.

The concept of domain specificity in its broadest application refers to the relationship between the various domains within which humans typically show expertise (such as face recognition or reading) and the cognitive processes and brain mechanisms that subserve these abilities. In the most colloquial (and inaccurate) version of this concept, there are brain structures 'for' face recognition or 'for' reading. This idea has been so influential (even when it was not explicitly articulated) that it has spawned separate research disciplines: visual neuroscience, cognitive

neuroscience and social neuroscience, for example, are all defined by their different domains and, secondarily, by a focus on different parts of the brain[1]. But is there really a visual brain, a cognitive brain and a social brain[2–4] (BOX 1)? These questions remain hotly debated particularly within social neuroscience.

Given the ever-increasing volume of data, especially those derived from neuroimaging studies of the human brain, we believe that it is timely to take stock of these questions. In this article, we aim to formulate a revised concept of domain specificity that draws lessons from the debates and that incorporates important processing features discovered within social neuroscience. The new view is more nuanced, accommodates a richer range of processing and should better guide future studies of the underlying neural circuits.

Origins of domain specificity

The first clear usage of the term domain specificity was provided by Jerry Fodor, who listed it as one of several criteria that described what he referred to as 'information processing modules' in a cognitive architecture[5]. In Fodor's view, these modules could be considered as little 'boxes' that encompassed the processing of the inputs to a general (non-modular) cognitive

system (FIG. 1). The modules had several attributes in addition to domain specificity, such as innateness (the modules placed bounds on what could be learned and ensured some universality in how different people's minds worked), informational encapsulation (the modules received restricted sensory information) and cognitive impenetrability (the modules could not be influenced by top-down cognition) (BOX 2). Informational encapsulation and cognitive impenetrability together limit the range of information to which a module has access (FIG. 1a). There is now general agreement that the original concept of modularity (or at least most readers' interpretation of it) needs to be made less rigid and to admit degrees of modularity[6]. However, there is also agreement that, although we may want to soften our adherence to the original list of specific criteria, a revised view of modularity remains powerfully useful in cognitive psychology and cognitive neuroscience. This is the view that we also take here: domain specificity is not exactly what we thought it was, but it is here to stay.

Of the original features in Fodor's list, one has been argued to be essential for modularity: domain specificity[7]. It is important to note that, in Fodor's original scheme, domain specificity arose from the joint effects of informational encapsulation and cognitive impenetrability; however, the strict operation of these two criteria depended on a particular concept of processing architecture that is incompatible with what we now know about the brain (FIG. 1). In particular, actual anatomical circuitry is now known to be too promiscuous to generate domain specificity in this way. Instead, we need a revised functional concept of domain specificity that would let us understand how a more realistic processing architecture (FIG. 1c) could still generate (degrees of) modularity.

In this regard, it is also important to distinguish between the inputs to a system (both 'bottom-up' and 'top-down') and the actual internal computations that the system is able to carry out, whatever the inputs. Both will constrain the outputs of the system: it can only operate on what it has access to, and it can only execute those computations that can be implemented in

its internal architecture. This is a distinction originally noted by Barrett[8] (a point that he illustrated with respect to the computations that enzymes carry out) who argued that either restricted access to information or restricted operations on that information (access versus process specificity, in Barrett's terms) could produce domain-specific outputs. Unlike Fodor, who restricted domain specificity to the peripheral modules that provided the inputs to a domain-general central system (FIG. 1), several recent thinkers have therefore advanced the view that even central processing could be domain specific[9,10]. Indeed, the 'massive modularity' hypothesis holds that central processing not only can, but must, be entirely domain specific: that the mind is composed of modules through and through[11].

The arguments supporting this idea include appeals to how psychological processes might have evolved, to computational efficiency and to the apparent requirement for innately specified heuristics (simple decision rules and problem-solving strategies) to make any aspect of cognition tractable.

Nowadays, the idea of domain specificity has crept into a large number of different studies that include evolutionary, developmental and clinical work[12–14]. Almost none of this work articulates exactly what the authors mean by 'domain specificity'. We believe that it will be important to consider the various difficulties facing the definition and assessment of domain specificity to forge a way forward towards a more nuanced concept that does not require all of the

commitments of Fodor's original definition. Although domain specificity was originally an abstract concept applied to any cognitive system, the revised concept will also need to incorporate actual data from neuroscience.

Continuing challenges

Conceptual problems. Confusions about domain specificity often arise as a result of dichotomous thinking about brain–domain relationships (domain specific versus domain general) and in specifying what one takes to be a domain (for example, social versus non-social, emotion versus cognition (BOX 1)). Moreover, debates have often ignored the fact that it is largely unknown how cognitive functions are related to brain structure and function[15,16]. Indeed, studies often consider only the
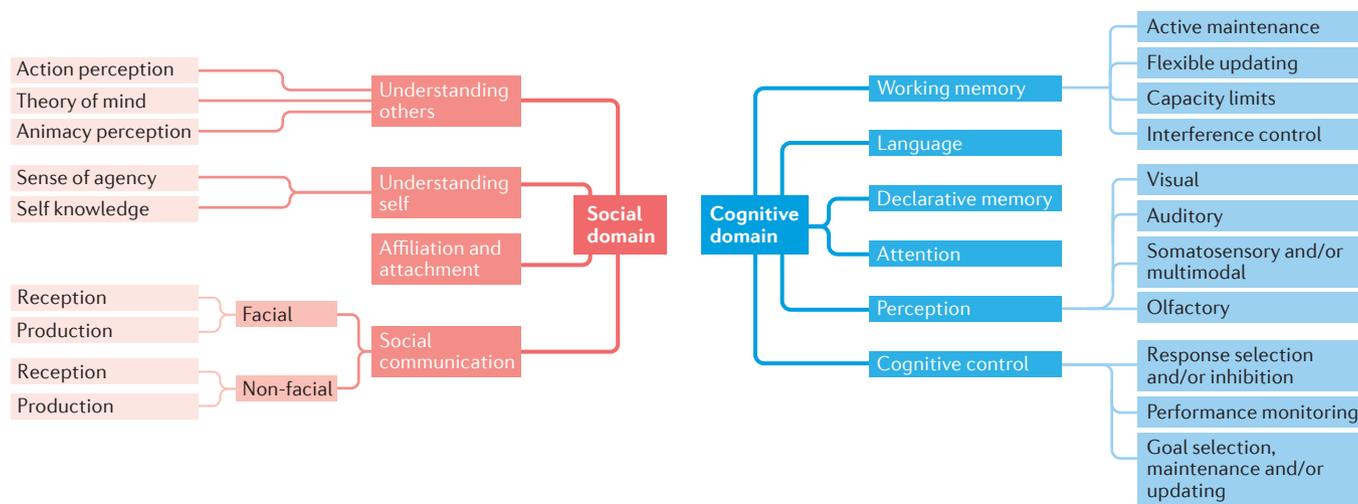

Box 1 | What is a cognitive domain?

In its most general sense, the domain of a cognitive function is the set of inputs and outputs to which it can be applied[5,69,70]; however, this quickly becomes complicated by the fact that domain specificity is nowadays commonly applied not only to cognitive processes but also (or instead) to brain regions or systems. Thus, face perception is a cognitive function for which there is substantial evidence for domain specificity, whereas the fusiform face area (FFA; see the main text and BOX 4) is one component of a distributed set of brain regions that generate this domain specificity through an assembly of computations.

For claims about domain-specific cognitive processes to be testable, there must be observable criteria that would allow us to judge whether a given set of inputs and outputs belong to the domain in question[71]. However, domains differ substantially in how easily such criteria can be defined. Specific classes of stimulus input (such as faces versus non-faces or words versus non-words) and motor output (such as eye movements or hand actions) are examples of relatively well-defined domains. Yet, as one moves further away from such concrete domains of stimulus and response, the inclusion criteria quickly begin to lose their definition (that is, there are less well-defined boundaries between domains). At the extreme is the distinction between the 'social' and 'non-social' domain. The social domain has characteristic stimuli (including faces, bodies and biological motion) and concepts (such as mental states); however, these

characteristics fail to deliver a coherent domain definition because humans can objectify social stimuli and can impute mental states to non-social stimuli.

One of the problems is that domain-based descriptions of brain and behaviour can vary substantially in their level of abstraction. To illustrate the challenge that this presents, as well as a potential solution, we can consider the 'social' and 'cognitive' research domains that were defined by the National Institutes of Mental Health Research Domain Criteria (RDoC) initiative[72–74] (see the figure). As exemplified in the RDoC initiative, these two very abstract domains are theoretically important; however, they must be decomposed into scientifically tractable subdomains that are specific enough to be operationally defined in the context of an experimental paradigm and that can be tied to specific measures of cognitive and neural function.

For claims about the domain specificity of neural structures or systems, analogous challenges arise, although the functional criteria of inputs (stimuli and/or tasks) and outputs (behavioural performance) are, in this case, importantly supplemented by neural measures such as electrophysiology or functional MRI. It is these additional measures, which arise from cognitive neuroscience, that can help to constrain the level of functional abstraction at which we explain domain specificity and that can provide a mechanistic explanation of how domain specificity at higher levels of abstraction emerges from the operation of lower-level mechanisms.
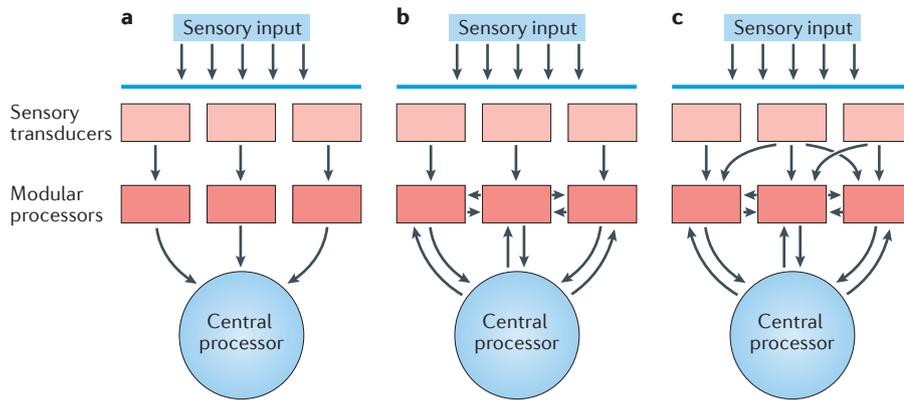
Figure 1 | **Fodor's conception of modularity versus 'new-look' modularity.** In the early 1980s, the philosopher Jerry Fodor provided the seminal framework for thinking about the mind as including independently operating functional units, or modules[5]. **a** | A Fodorian module is shown. According to Fodor's original concept, modular input systems process information in a domain-specific manner, because they only have access to one specific channel of input (the domain on which they could operate), a concept known as informational encapsulation. That is, each modular processor receives information from a dedicated set of sensory transducers. In addition, Fodor proposed that there is a central processor that receives information from the modular processors and is domain general, because it has access to everything (or at least to a flexibly unbounded number of things). **b** | According to Fodor, violations of informational encapsulation as a result of reciprocal inputs between processors would destroy modularity, because it would destroy domain specificity. **c** | In our 'new-look' version of modularity, domain specificity can emerge if a module has so much access to information (because it receives information from many sensory transducers together with top-down modulation) that it is able to intelligently filter that information to yield what looks like central domain specificity.

small subset of anatomical loci for which standardized (univariate) neuroimaging localizer protocols are available[17,18] or for which there exist rare patients with anatomically circumscribed lesions. Such scattered and selective evidence for domain specificity does not allow us to discover the full architecture of cognition and may be merely imposing preconceived views that are perpetuated by selecting among the scattered reports. Moving forward, the concept of domain specificity will need to incorporate the accumulating evidence that 'natural' domains — that is, the domains gleaned from experiments in a data-driven manner — are in many cases going to be more abstract than our common-sense psychology might have us intuit and are in turn likely to correspond to a dynamic, network-based picture of brain function that is often also counterintuitive[19–21].

Empirical problems. These conceptual difficulties are closely related to problems in how we collect empirical data and to the inferential logic whereby neuroscientific evidence is used to promote claims of domain specificity. The most blatant of these problems is the over-reliance on data demonstrating merely that a region (such as the dorsomedial prefrontal cortex (dmPFC))

shows a greater functional MRI (fMRI) response (commonly termed an 'activation') to stimuli that are exemplars of one domain (such as human behaviours) than it does to stimuli that are exemplars of another domain (such as non-human natural phenomena). At best, such data demonstrate domain specificity in the response, and, in most cases, they can be easily explained as the effect of a range of third variables — some of which may be theoretically interesting in their own right (such as the reward value or familiarity of the stimuli). According to our view, it is insufficient to demonstrate domain specificity in a brain region without some explanation of why that specificity exists. After all, any domains of stimuli that we can tell apart would be expected to be represented differently in the brain at some level: if this were not the case, on what basis could the domains be discriminable by people in the first place?

As mentioned above, the concept of domain specificity begs the question of what counts as a domain (BOX 1). In a trivial sense, a domain can, of course, be specified just in terms of stimulus input; however, even here, issues arise owing to our inability to exhaustively search for responses to all stimuli, and there may be problems in eliminating confounds (as noted above).

In addition, we are usually not interested merely in stimuli and tasks but rather in latent (not observable) variables that can only be inferred from those stimuli and tasks. This issue is particularly acute in social neuroscience, in which different investigators (and studies by the same investigators) often use different stimuli and tasks to putatively measure the same process. A clear example of this (which we discuss further below) comes from the hundreds of human neuroimaging studies of 'theory of mind' (ToM), the ability to impute mental states to other agents. There is enormous diversity in the tasks and methods that are used to assess ToM, including simple lexical decisions about mental-state verbs, passive viewing of geometric animations and playing a competitive economic game with another person[22]. The assumption that there is a single ToM ability underlying performance in such diverse tasks has been largely untested, and recent empirical evidence strongly suggests it is untenable[17,23,24]. In short, this leads to important questions about construct validity: a single finding of a response that seems to be domain specific is almost certainly inadequate. Instead, we would want an accumulation of evidence (ideally multimodal) that together provides convergent validity: that is, a number of different measures that all provide evidence that a particular variable shows domain specificity in its processing.

Mechanisms. An important ongoing debate concerns the mechanisms that could underlie domain specificity. As noted above, domain specificity may theoretically arise either because the processing 'module' is connected in such a way that it only receives restricted information ('input specificity' in our terminology) or because its internal computations only accept a certain range of information ('central specificity' in our terminology). One might consider the visual system as an example of the former mechanism, whereas language is an example of the latter.

It is important to note that, in both cases, the putative domain specificity must depend on architectural processing constraints and not on a host of other factors that could produce illusory domain specificity and that are typically eliminated in control experiments, such as one's preferences or cognitive limitations[25]. For example, I may not be able to process and perform arithmetic with very long numbers that exceed the capacity of my working memory, but this doesn't mean that there is domain specificity for short

Box 2 | **Domain specificity in the broader context of cognitive architecture**

In Fodor's seminal text[5], domain specificity is presented as one of several design features to consider in descriptions of cognitive architectures (models of how the mind works that range from classical 'boxes-and-arrows' schemes to detailed connectionist networks inspired by models of neurons). As stressed by Fodor himself (and others since[7]), these proposed features were not intended to provide a definitive set of conditions that are necessary and sufficient for a modular design and would not always be expected to co-occur. We therefore believe that future research on neural architecture would benefit from treating each of these features as an empirical issue unto itself. In this spirit, we have placed each of Fodor's original features, along with others that have since emerged, into one of four categories (see the figure). This is not meant to be an exhaustive feature list; rather, it illustrates that domain specificity is just one small piece of a much larger puzzle.

Implementation
Recent data from studies examining intrinsic functional architecture in the resting brain suggest that characteristics of modular design exist at the level of large-scale networks of spatially distributed and functionally connected brain regions[20,21]. Whereas Fodor emphasized innateness, modern studies have also highlighted the role of experience-dependent functional plasticity[75,76] and, relatedly, variability across individuals in function-to-structure mappings[77].

Deployment
Fodor expected that modular functions would deploy efficiently and automatically and would be cognitively impenetrable: that is, they would neither require nor permit intentional or conscious control over when and how they were deployed.

Function
Although questions of domain specificity can be tackled in the absence of a precise account of the mechanisms through which that domain specificity is generated, we suggest that it is important to understand the latter to explain the former. This might be especially the case for truly modular functions, which Fodor argued would be likely to be computationally simple (non-decomposable), as the corresponding causal mechanisms should be simpler to discover.
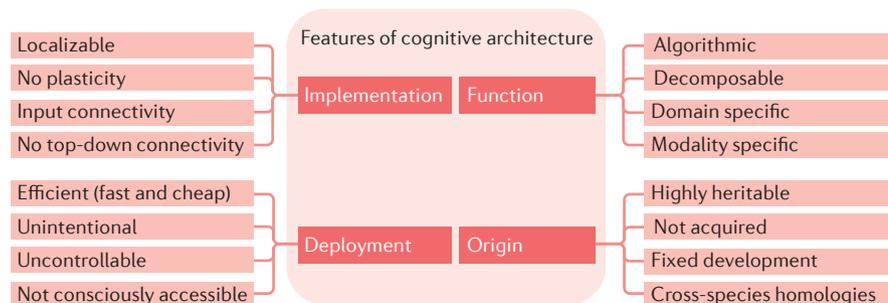
Origin
According to some accounts[9,10], modules are evolved adaptations that enabled us to solve specific, recurring problems in our ancestral environments. Thus, this feature set is tightly linked to those related to implementation: purely modular functions have high heritability, low environmental influences, a characteristic developmental trajectory and are likely to have cross-species homologies.

Implementation	Function
Localizable	Algorithmic
No plasticity	Decomposable
Input connectivity	Domain specific
No top-down connectivity	Modality specific

Features of cognitive architecture

Deployment	Origin
Efficient (fast and cheap)	Highly heritable
Unintentional	Not acquired
Uncontrollable	Fixed development
Not consciously accessible	Cross-species homologies

numbers and longer numbers. Similarly, I may have a strong preference for (and hence attend to and remember things about) Golden Retrievers, but this doesn't mean that I have domain-specific processing for Golden Retrievers. One could certainly imagine that I might, over time, develop domain specificity as a result of these other influences; however, again, this would have to manifest at the level of changes in the processing architecture. Data from neuroscience can sharpen our understanding of the processing architecture and can therefore provide some of the most compelling arguments for particular mechanisms.

There are, in fact, neurobiological data that support the idea that central specificity can develop over time. For example, histological data show that cortical proliferation changes during development in regions that subserve face processing[26]. Such ultrastructural changes provide good candidate mechanisms that may drive the emerging specialization in the intrinsic computations carried out within cortical regions that exhibit domain specificity. Although this does not of course eliminate the possibility that there is input specificity (which may, after all, also be the ultimate cause of the changes in intrinsic

specialization), these findings do argue for a view according to which domain specificity — at least for central processes — arises from a mixture of sources. Sensory input patterns, access to other information through specific patterns of connectivity and intrinsic operations that are evident in ultrastructural measures are all likely to contribute to the domain specificity of a system or region (BOX 3).

Development. Questions about development have traditionally been at the core of debates about both modularity (remember that one of Fodor's original criteria was innateness) and domain specificity. Waddington famously proposed an image in which a developmental trajectory could be represented as a marble rolling down a slope with valleys in it: the valleys buffered the developmental course against sources of variance and ensured that the process became 'canalized' (REF. 27). Several modern views have picked up this theme and stress that the specialization of a process during development is driven by a mixture of genetic and environmental mechanisms that are often experience dependent[28]. For instance, this has been applied to the much-debated domain specificity of 'mirror' neurons, which participate in both the production of actions and in the perception of similar actions in other agents: an infant's own movements, together with attentional preferences and Hebbian learning, can lead to the emergence of domain specificity in the responses of mirror neurons[29]. This example is particularly attractive, as it makes precise predictions about the parameters that are required for the emergence of domain specificity, based on what is known about Hebbian learning. It is also a clear example of a specifically social process that is rendered domain specific through canalization of a previously domain-general process.

There is a final important way in which originally domain-general processing might become more domain specific through experience: the operation of interpersonal factors. The inputs to a system are dictated not only by the preferences of the individual and the statistical structure of the world but are also actively shaped by social and cultural constraints as multiple people interact with one another[30]. Indeed, one would conjecture that similarities across individuals in the degree to which many social processes are domain specific may well arise from the coupled social

Box 3 | Mapping domain specificity in the brain

We suggest that there are three broad classes of architectures that might generate domain specificity (not mutually exclusive). The first would arise from the restriction of sensory inputs to the domain (input specificity), the second is the result of the specificity of computations internal to a central module (central specificity), and the third arises from the existence of horizontal and top-down inputs to a system that act as sophisticated filters (new-look specificity) (FIG. 1a,b,c, respectively). How and where might we find evidence for each of these types of processing architecture in the brain?

   We could very roughly partition the brain into early sensory regions, sensory association cortices and central regions (excluding the motor-related structures that relate to output for now). In early sensory structures, we might expect to find evidence primarily for input specificity (as evident in the vomeronasal system (see the main text)). The topographic maps that characterize early sensory regions also provide evidence consistent with input specificity (for example, the tonotopic auditory cortex could trivially be said to be domain specific for a particular frequency). However, it would be more interesting if we could find topography that maps semantic dimensions. It is possible that these could be found in association cortices, where they would need to be centrally synthesized. Examples of such findings include evidence of cortical specializations for particular object categories (including the fusiform face area (FFA) (BOX 4)) and their lexical representations, such as maps representing semantic and lexical information about people, animals, fruits, tools and musical instruments[78,79]. What remains debated is how these central representations arise. According to our view, they are likely to arise as a result of new-look specificity: they would depend on the complex patterns of connectivity that enable a region of the brain to recruit the many associations that constitute our knowledge about these different domains of concrete entities.

   More 'central' regions, such as the dorsomedial prefrontal cortex (dmPFC), seem to exhibit relatively domain-general processing when appropriate control conditions are included[56] or when a number of ostensibly non-social domains are being examined[42]. There is little argument for input specificity in these regions, and the apparent domain specificity of these areas is so flexibly deployed that it seems to be truly dynamical: that is, it is engaged in the moment and subject to momentary motivational and attentional effects[80–82]. It may be that these effects result in some enhancement of processing within the dmPFC. This could involve neuronal proliferation (as is the case for the FFA[26]) or other plastic events within the dmPFC. Although normal (domain-specific) experience during development may thus be one factor driving maturation of the dmPFC, the computations that it performs would nonetheless remain inherently domain general. It is possible that there are other 'central' regions that do exhibit domain specificity to some degree, such as the adjacent regions in the medial PFC that seem to represent value for rewards such as food or money[83].

   We are left then with clear input specificity only for sensory regions. As noted above, domain specificity in association and higher-order regions is likely to result from new-look specificity. There is little evidence of true central specificity, which either turns out to be new-look specificity or else reflects domain-specific developmental or experiential effects on the functioning of an otherwise domain-general process.

interactions between individuals. The need for efficient interpersonal communication and the creation of substrates for shared experience through culture could drive the emergence of domain specificity for many social abilities, such as reading text[31].

Social neuroscience case studies

All information is conveyed to the brain through inputs that ultimately depend on sensory transduction. Once in the brain, there is only a single, intrinsically content-less (and so completely domain-general) currency: electrical potentials. Given this fact, the ultimate source of domain specificity must be either in the biophysical constraints on sensory transduction channels (and their subsequent pattern of disseminating the transduced information; input specificity) or in the nature of the computations carried out by
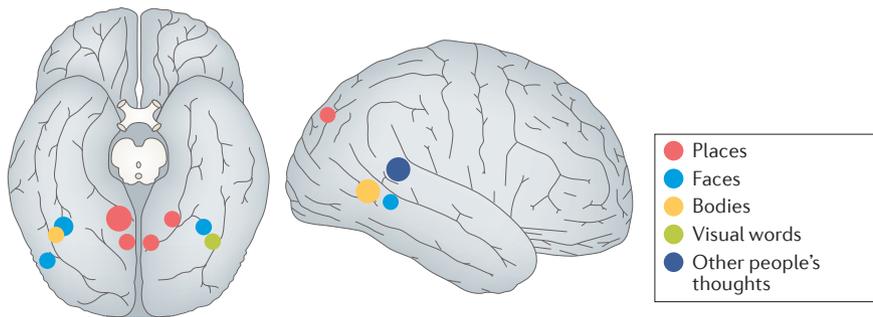
a central processor (central specificity), or both. We can gain insight into which of these possibilities is the more plausible in different scenarios by considering several examples of putative domain specificity in social neuroscience studies.

   As an example of input specificity, we can consider the social behaviours that are evoked by pheromones in many mammals and insects. These are caused by specific molecules and are processed by a neural system (the vomeronasal system) that meets many of the classic criteria for a cognitive module: it is sensitive only to the particular pheromone molecules and it links its processing narrowly to a specific class of behaviours (BOX 2). The mating behaviour elicited in the males of many species by the smell of females in oestrus, for instance, is an example of domain-specific processing that arises at the level of sensory transduction[32,33].

Nonetheless, even here, it is apparent that such domain-specific processing is merely one component of a much more complex architecture that ultimately generates flexible behaviour in context. For instance, domain-specific input modules can still engage central states that are more flexible, even for brains as 'simple' as that of the honeybee[34].

   Another example of putative domain specificity comes from face processing (BOX 4). Neuroimaging studies have shown that a region of the ventral temporal cortex, dubbed the fusiform face area (FFA), is activated more by faces than by any other domain of visual stimuli[35] and contains neurons that respond selectively to faces[36]. In this case, what seems to be domain specificity for a certain kind of input must in fact arise from domain specificity due to certain kinds of computations, as there are no inputs (at the periphery) that are selective for faces. The explanation of how domain specificity is centrally generated in the case of face processing may be a complex account of successive filtering and combinatorial operations that, in and of themselves, do not operate only on input about faces (as, again, there is no such selectivity in terms of sensory inputs). Considerable debate about the domain specificity of face processing has focused on the question of whether the FFA responds selectively to faces or instead implements a computational function that happens to be most important for processing faces (BOX 4). In fact, these may simply be accounts of two different things: there is domain-specific processing for faces that arises because of central computations that construct such processing. These central computations are now being revealed in a network of patches of temporal cortex that each contributes to the components required[37–39].

   Central domain selectivity can thus arise not simply from the input connectivity that conveys sensory information as such but from the full pattern of connectivity through which a central brain region obtains highly transformed and context-modulated sensory information. The idea that a particular connectivity profile could form the substrate for the domain-specific function of a brain region is borne out by recent findings about word processing. Unlike face processing, it seems that reading words is a capacity that has emerged too recently for us to have evolved a module that is domain specific for words. However, recent work has shown that the structural connectivity of the region that normally develops into the visual

---

Box 4 | Domain-specific functions in temporal cortex

A large number of studies of primate brain function have identified several anatomically specific regions of the temporal cortex that seem to be specialized for processing a single category of stimuli[63,65,84]. There are regions that respond most to faces, to places, to bodies or to visually presented words. Each of these areas has its own name: the fusiform face area (FFA), parahippocampal place area, extrastriate body area and the visual word form area, respectively. The approximate size and location of these regions are indicated by the size and position, respectively, of the circles in the figure. We also depict a region at the temporoparietal junction (TPJ) that is thought to be specialized for theory of mind (the ability to represent other people's thoughts)[46,85]. Although the evidence that each of these regions has a 'preferred' domain is practically indisputable, the extent to which the function implemented in these regions is domain specific is a matter of continuing debate.

Although a complete account of the case against domain specificity in the temporal cortex is outside the scope of the present article, this argument has largely centred on evidence demonstrating that the selectivity of the FFA to exemplars from non-face domains (such as butterflies or cars) tracks individual differences in domain-specific perceptual expertise[86–88]. This has been suggested to imply that the FFA acquires its selectivity for the face domain not because it performs a function that is specific to that domain but because the function it performs enables the early acquisition of a fundamental social skill: recognizing conspecifics based on subtle variation in facial appearance[53,89]. A similar expertise-based account has been advanced to explain the selective responses to mental-state reasoning in regions such as the TPJ and (as discussed in the main text) the dorsomedial prefrontal cortex[31,55,56,90]. However, advocates of the domain-specificity account of the temporal cortex claim that learning mechanisms cannot explain all the available evidence[91], noting the face-specific recognition deficits seen in prosopagnosia[92] and recent evidence demonstrating that face-selectivity temporal cortex is already apparent in 4-month old infants[12]. Therefore, despite the more than two decades of research and discussion that have been generated following the publication of the original neuroimaging findings in this area, there is still no clear consensus regarding the domain specificity of functions in the temporal cortex. Figure adapted with permission from: Kanwisher, N., Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl Acad. Sci. USA* **107**, 11163–11170 (2010).

Places
Faces
Bodies
Visual words
Other people's thoughts

---

word form area (VWFA) (BOX 4) is already preconfigured to become specialized for processing words given the right subsequent experience[40]. Indeed, the precise anatomical location of what will later become the functionally defined VWFA can be determined 3 years earlier in infants (before reading competency has been acquired) from the structural connectivity profile of this region. This suggests one mechanism by which central specificity might arise: domain-specific computations within a brain region or system might emerge when particular connectivity meets experience with a particular domain. In the case of reading, the VWFA is not domain specific for words at birth but progressively becomes so as a particular processing architecture (that is, a particular pattern of connectivity) is applied to a restricted content over time.

This brings us to our final example, the putative domain specificity of social cognition more generally. Whereas the FFA and VWFA are high-level perceptual regions (association cortices) (BOX 3), thinking and reasoning about other people's minds is a cognitive ability that can operate on a huge range of stimulus inputs and relies on a network of very central regions (such as the dmPFC)[22,24,41,42]. Thus, Fodor would have put it squarely into the domain-general 'central system' box (FIG. 1). However, the idea that social information processing is domain specific has a long history, going back to some of the origins of social psychology[43] and the notion of a 'social brain' (REFS 3,4). For instance, lesions to certain brain regions — such as the ventromedial PFC or amygdala — often result in disproportionate impairments in social cognition, and

differential processing of social information has often been highlighted in disorders such as autism and Williams syndrome[44]. Neuroimaging data broadly support the dissociation between social and non-social cognition[1], and there is even evidence that human children exhibit cognitive skills notably superior to those of great apes specifically in the social domain[45]. As in the case of reading, social cognition and its neural substrates have thus been proposed as a type of domain specificity that is possibly unique to humans[46].

More recent work has focused on a network of structures that are involved in our ability to represent and reason about other people's mental states. As noted above, this is often dubbed ToM, although that term is applied to several different kinds of social reasoning that each recruits somewhat different brain regions[17,23,24]. Perhaps the most prominent of these are two nodes in this network: the dmPFC and the temporoparietal junction. The dmPFC in turn can be further fractionated into subdivisions[41,47]. Hundreds of neuroimaging studies of higher-order social cognition have led many to conclude that the dmPFC performs a function that is dedicated to the decidedly social task of reasoning about the mental states of others[22,47,48]. Moreover, recent neuroimaging studies of the macaque suggest that several cortical regions — including most prominently regions of the PFC believed to be homologous to the human dmPFC — are specialized for interpreting the contents of complex social scenes[49], such as the information extracted from the faces of conspecifics[50]. These studies, along with related proposals from human neuroimaging studies[51], indicate that specialization for social cognition may be best characterized at the level of functional circuits consisting of multiple components that themselves need not be domain specific[52].

Although a role for the dmPFC in ToM is practically indisputable, there are reasons to question the stronger claim that the dmPFC is domain specific for social processing. Most problematic is the fact that the neuroimaging evidence for domain specificity seems to rest, almost exclusively, on showing that the dmPFC responds more strongly under social conditions than under non-social conditions[1]. Such a differential response could easily be explained by the existence of a single, domain-general process that is deployed to a greater degree in the social domain because of a range of possible third variables that confound the

social–non-social comparison, a common confounding problem that we noted earlier. For instance, greater deployment of the dmPFC in the social domain could be explained by the acquisition of greater levels of domain-specific expertise during typical development[53–55]. Further supporting a more domain-general account of the role that the dmPFC has in ToM, a growing body of neuroimaging studies suggests that it plays a functional part in a range of cognitive domains that do not fit squarely into the domain of social cognition. These include causal explanations of natural phenomena[56], idiom comprehension[57], acquiring conceptual understanding of novel tools[58], selecting among multiple semantic memories[59], mentally simulating and evaluating novel future outcomes based on simultaneously activating memories of multiple past outcomes[60], and acquiring and using abstract (non-spatial) conceptual knowledge[61,62]. As with the role of the FFA in face processing, finding domain-specific social functions in the dmPFC may thus be explained by the application of these multiple contributing components to a particular class of stimuli.

We therefore believe that the dmPFC implements a function that is crucial to the acquisition of domain-specific conceptual expertise (whether social or non-social) and to the use of such expertise to explain and understand the events that we observe in the world (whether social or non-social). This conclusion is further supported by examining individual differences: the people who are the most curious and interested in others are also those who show disproportionately greater activation of the dmPFC in response to social stimuli[56]. As is the case for reading and the VWFA, and as is the case for face perception and the FFA, it seems to be the interaction of a broader computational function with particular experience-dependent expertise that yields domain specificity.

Domain specificity revised

Despite the noted debates about the domain specificity of regions such as the FFA, there are strong arguments to support the view that face processing is domain specific by subject matter; that is, it is not just a result of particular functions like expert within-class categorization but is indeed specialized for faces per se[63–65] (BOX 4). But what could this possibly mean? In what way could a central process be specialized for a certain class of stimuli, if not either by restricted inputs or by specialized computations? As mentioned briefly above, one possible

explanation considers the importance of the connectivity profile of a region together with experience in a specific domain. This leads to the following hypothesis: central domain specificity could result if, in addition to certain computations internal to a region (which could be applied more domain-generally in principle, as is in fact the case for the FFA just as it is for the dmPFC), there is also modulation by many other sources of input (including 'top-down' effects). In this case, the process that is under consideration would, in a sense, 'know' that the stimuli are faces or that the reasoning is social and hence process the information that it receives differentially based on such contextual factors. After all, you yourself know that a stimulus is a face or an inference is social, and you can easily treat these in a domain-specific manner if you so choose. If a 'module' for face processing had that kind of broad information available, it could also treat a face as different from a non-face — not because it only gets information about faces, but rather because it gets so much information that it can decide to process faces differently from non-faces. By allowing higher-order beliefs and goals to have an influence on ('penetrate') the internal operation of a module, this new version of domain specificity violates one of Fodor's original criteria[5] (BOX 2) for a module: that its operation be cognitively impenetrable. We concur with others[7] that domain specificity should be the defining characteristic of a module; however, we go further in suggesting that informational encapsulation and cognitive impenetrability are not only unnecessary but in fact counter-productive for producing domain specificity. We suggest that a high degree of domain specificity can in fact arise from a marked absence of information restriction: the more information a face-processing module has available, the more sophisticated the neural decision can be to process faces differently from non-faces.

This 'new look' version of domain specificity may sound like we are putting a homunculus inside the module, but we are merely noting that apparent specialization for a domain can emerge just because all knowledge about a domain is ultimately represented in the associations that we have stored in our brain. If facts about a domain that are acquired through experience are encoded in patterns of synaptic weights, as we know they must be, we can think of these as filters operating on inputs to, and modulating the internal operations of, a module. In their aggregate operation, these

filters may seem as intelligent as the person; however, we see no contradiction here. We may detect this version of domain specificity through the processing of the module, but this processing may in fact reflect the operation of the rest of the brain. This idea becomes more concrete when we remind ourselves of the main contributions to the blood-oxygen-level-dependent (BOLD) signal typically measured in fMRI studies: it does not reflect actual spiking outputs from a region but rather synaptic inputs to, as well as intrinsic processing within, the system[66]. It is thus quite conceivable that a neuroimaging study that detects domain-specific activation in a brain region is in fact detecting not so much the domain specificity of outputs from that region but rather the joint input function that is specified by all the synapses made onto this region from elsewhere in the brain.

In emphasizing a more dynamic view of domain specificity, three different scales of possible change should be considered: phylogenetic, ontogenetic and truly dynamic (real-time, momentary modulation by attention and context). All these changes can influence processing architecture and hence lead to the emergence of domain specificity, and they do so in an approximately hierarchical manner: evolution has sculpted some architectural limits on plasticity and reorganization; within this scope, development and experience can reconfigure how information flows in the brain; and within this scope, dynamic routing can change even the current processing architecture functionally. The latter case is closely aligned with modern concepts of 'neural reuse' (REF. 67) and 'cultural recycling' (REF. 68) and indeed does challenge the original definition of domain specificity that we sketched at the very outset. If domain specificity can arise, perhaps transiently, from the full flexibility of information routing in the brain, it looks just like the confounding effects of attention or preference that one would like to eliminate (I might end up with a module for Golden Retrievers after all).

In our view, central domain specificity is best conceived as one of many dimensions on which mental functions might vary (BOX 2). It may arise from partly innate factors and develop early, as seems to be the case for face processing. Alternatively, it may also arise later and develop as a result of very specific (and perhaps culturally transmitted) kinds of experience, as seems to be the case for reading. Finally, it may arise throughout the lifespan given sufficient structure in

motivation, attention or perception, as seems to be the case for social inferences that rely on the dmPFC. The confounding case of merely thinking differently about one domain compared with another (owing to momentary differences in attention or preference) would be at the extreme end of this spectrum and would only qualify as the weakest form of domain specificity if it becomes so systematic that it begins to characterize an architectural feature of that person's cognition. Over time, of course, such weak domain specificity can become canalized into stronger forms, as (in our view) is the case with social inference and the dmPFC. Thus, domain specificity is certainly not black and white but is rather continuous, and even the faintest versions of it can, over time, explain how the stronger versions emerge.

Concluding remarks

Neuroscience data will be essential to distinguish between and provide evidence for the different types of domain specificity that we propose. It may also provide an account that locates the particular case under investigation on a continuum of input specificity to central specificity and that explains how the latter is implemented as the result of connectivity and experience. One type of such data that will be particularly crucial to the revised view of domain specificity that we have sketched is developmental and longitudinal data that can quantify how information processing changes within an individual. This is a type of data of which we have little, because it is difficult to collect. Future studies should also consider data-driven discovery of domain specificity, as well as modelling how domain specificity might spontaneously emerge in synthetic systems; the latter might provide an important complement to the collection of dense longitudinal data in humans. Finally, an overarching question is normative: which aspects of domain specificity might constitute the core of typically developing human cognition? Is there a collection of architecturally constrained processes, such as those for language, faces and social inference that together make human cognition different from the cognition of other animals?

Our new view of how domain specificity can in fact arise through the violation of cognitive impenetrability suggests that, to understand any component of mental architecture, we might need to understand all of it. Fodor already acknowledged that this kind of holism was likely to be true and concluded that very little about cognition could in fact be understood through the operation of classical modules[25]. Whereas this led Fodor to a jaundiced view of our prospects for ever understanding the mind, we take a more cheerful attitude because we include cognitive neuroscience in our data. This source of data helps to provide constraints because it can show us the actual mechanisms whereby a module receives the inputs that bestow it with domain specificity.

The concept of domain specificity had a huge role in the formative years of the cognitive neurosciences. It is our view that the basic questions around which it revolves and the theoretical concepts it is founded upon will continue to provide fertile ground moving forward. However — and to finally make the obvious pun — we need to be more specific in how we conceive and empirically investigate not just those domains that our minds evolved to solve but also those domains that (through the accumulation of culture, history and technology) present our minds with endless domains in which to test its limits. A new view of domain specificity emphasizes both the limits and the extensions of human cognition.

Robert P. Spunt and Ralph Adolphs are at the California Institute of Technology, 1200 E California Blvd, HSS 228–77, Pasadena, California 91125, USA.

radolphs@caltech.edu; spunt@caltech.edu

1. Van Overwalle, F. A dissociation between social mentalizing and general reasoning. *Neuroimage* **54**, 1589–1599 (2011).
2. Adolphs, R. The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* **60**, 693–716 (2009).
3. Brothers, L. The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51 (1990).
4. Dunbar, R. I. The social brain hypothesis. *Evol. Anthropol.* **9**, 178–190 (1998).
5. Fodor, J. A. *The Modularity of Mind: An Essay on Faculty Psychology* (MIT Press, 1983).
6. Barrett, H. C. & Kurzban, R. Modularity in cognition: framing the debate. *Psychol. Rev.* **113**, 628–647 (2006).
7. Coltheart, M. Modularity and cognition. *Trends Cogn. Sci.* **3**, 115–120 (1999).
8. Barrett, H. C. Enzymatic computation and cognitive modularity. *Mind Language* **20**, 259–287 (2005).
9. Barkow, J. H., Cosmides, L. & Tooby, J. (eds) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford Univ. Press, 1995).
10. Pinker, S. How the mind works. *Ann. NY Acad. Sci.* **882**, 119–127 (1999).
11. Carruthers, P. in *Contemporary Debates in Cognitive Science* (ed. Stainton, R.) 205–225 (Blackwell, 2005).
12. Deen, B. *et al.* Organization of high-level visual cortex in human infants. *Nat. Commun.* **8**, 13995 (2017).
13. Kovacs, A. M., Teglas, E. & Endress, A. D. The social sense: susceptibility to others' beliefs in human infants and adults. *Science* **330**, 1830–1834 (2010).
14. Weiner, K. S. & Grill-Spector, K. The evolution of face processing networks. *Trends Cogn. Sci.* **19**, 240–241 (2015).
15. Poldrack, R. A. & Yarkoni, T. From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* **67**, 587–612 (2016).
16. Poldrack, R. A. Can cognitive processes be inferred from neuroimaging data. *Trends Cogn. Sci.* **10**, 59–63 (2006).
17. Spunt, R. P. & Adolphs, R. Validating the why/how contrast for functional mri studies of theory of mind. *Neuroimage* **99**, 301–311 (2014).
18. Saxe, R. Why and how to study theory of mind with fmri. *Brain Res.* **1079**, 57–65 (2006).
19. Sporns, O. Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* **17**, 652–660 (2014).
20. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
21. Sporns, O. & Betzel, R. F. Modular brain networks. *Annu. Rev. Psychol.* **67**, 613–640 (2016).
22. Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* **48**, 564–584 (2009).
23. Schaafsma, S. M., Pfaff, D. W., Spunt, R. P. & Adolphs, R. Deconstructing and reconstructing theory of mind. *Trends Cogn. Sci.* **19**, 65–72 (2015).
24. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* **42**, 9–34 (2014).
25. Fodor, J. A. *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology* (MIT press, 2001).
26. Gomez, J. *et al.* Microstructural proliferation in human cortex is coupled with the development of face processing. *Science* **355**, 68–71 (2017).
27. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
28. Gottlieb, G. Experiential canalization of behavioral development: theory. *Dev. Psychol.* **27**, 4–13 (1991).
29. Del Giudice, M., Manera, V. & Keysers, C. Programmed to learn? The ontogeny of mirror neurons. *Dev. Sci.* **12**, 350–363 (2009).
30. Schilbach, L. *et al.* Toward a second-person neuroscience. *Behav. Brain Sci.* **36**, 393–414 (2013).
31. Heyes, C. M. & Frith, C. D. The cultural evolution of mind reading. *Science* **344**, 1243091 (2014).
32. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat. Rev. Neurosci.* **4**, 551–562 (2003).
33. Keverne, E. B. Importance of olfactory and vomeronasal systems for male sexual function. *Physiol. Behav.* **83**, 177–187 (2004).
34. Menzel, R. & Giurfa, M. Cognitive architecture of a mini-brain: the honeybee. *Trends Cogn. Sci.* **5**, 62–71 (2001).
35. Kanwisher, N., Mcdermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
36. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
37. Meyers, E. M., Borzello, M., Freiwald, W. A. & Tsao, D. Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. *J. Neurosci.* **35**, 7069–7081 (2015).
38. Afraz, A., Boyden, E. S. & Dicarlo, J. J. Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proc. Natl Acad. Sci. USA* **112**, 6730–6735 (2015).
39. Aparicio, P. L., Issa, E. B. & Dicarlo, J. J. Neurophysiological organization of the middle face patch in macaque inferior temporal cortex. *J. Neurosci.* **36**, 12729–12745 (2016).
40. Saygin, Z. M. *et al.* Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* **19**, 1250–1255 (2016).
41. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
42. Spreng, R. N., Mar, R. A. & Kim, A. S. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* **21**, 489–510 (2009).
43. Heider, F. *The Psychology of Interpersonal Relations* (John Wiley & Sons, 1958).
44. Kennedy, D. P. & Adolphs, R. The social brain in psychiatric and neurological disorders. *Trends Cogn. Sci.* **16**, 559–572 (2012).

45. Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B. & Tomasello, M. Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**, 1360–1366 (2007).
46. Saxe, R. Uniquely human social cognition. *Curr. Opin. Neurobiol.* **16**, 235–239 (2006).
47. De La Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* **36**, 6553–6562 (2016).
48. Mitchell, J. P. Contributions of functional neuroimaging to the study of social cognition. *Curr. Dir. Psychol. Sci.* **17**, 142–146 (2008).
49. Sliwa, J. & Freiwald, W. A. A dedicated network for social interaction processing in the primate brain. *Science* **356**, 745–749 (2017).
50. Schwiedrzik, C. M., Zarco, W., Everling, S. & Freiwald, W. A. Face patch resting state networks link face processing to social cognition. *PLoS Biol.* **13**, e1002245 (2015).
51. Spunt, R. P. & Lieberman, M. D. An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *Neuroimage* **59**, 3050–3059 (2012).
52. Cabeza, R. & Moscovitch, M. Memory systems, processing modes, and components: functional neuroimaging evidence. *Perspect. Psychol. Sci.* **8**, 49–55 (2013).
53. Bukach, C. M., Gauthier, I. & Tarr, M. J. Beyond faces and modularity: the power of an expertise framework. *Trends Cogn. Sci.* **10**, 159–166 (2006).
54. Lamba, S. & Mace, R. Demography and ecology drive variation in cooperation across human populations. *Proc. Natl Acad. Sci. USA* **108**, 14426–14430 (2011).
55. Korman, J., Voiklis, J. & Malle, B. F. The social life of cognition. *Cognition* **135**, 30–35 (2015).
56. Spunt, R. P. & Adolphs, R. Folk explanations of behavior: a specialized use of a domain-general mechanism. *Psychol. Sci.* **26**, 724–736 (2015).
57. Lauro, L. J., Tettamanti, M., Cappa, S. F. & Papagno, C. Idiom comprehension: a prefrontal task? *Cereb. Cortex* **18**, 162–170 (2008).
58. Menz, M. M., Blangero, A., Kunze, D. & Binkofski, F. Got it! Understanding the concept of a tool. *Neuroimage* **51**, 1438–1444 (2010).
59. Satpute, A. B., Badre, D. & Ochsner, K. N. Distinct regions of prefrontal cortex are associated with the controlled retrieval and selection of social information. *Cereb. Cortex* **24**, 1269–1277 (2013).
60. Barron, H. C., Dolan, R. J. & Behrens, T. E. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat. Neurosci.* **16**, 1492–1498 (2013).
61. Kumaran, D., Summerfield, J. J., Hassabis, D. & Maguire, E. A. Tracking the emergence of conceptual knowledge during human decision making. *Neuron* **63**, 889–901 (2009).
62. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016).
63. Kanwisher, N. Domain specificity in face perception. *Nat. Neurosci.* **3**, 759–763 (2000).
64. Kanwisher, N. & Yovel, G. The fusiform face area: a cortical region specialized for the perception of faces. *Phil. Trans. R. Soc. B* **361**, 2109–2128 (2006).
65. Tsao, D. Y. & Livingstone, M. S. Mechanisms of face perception. *Annu. Rev. Neurosci.* **31**, 411–437 (2008).
66. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fmri signal. *Nature* **412**, 150–157 (2001).
67. Anderson, M. L. Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* **33**, 245–266 (2010).
68. Dehaene, S. & Cohen, L. Cultural recycling of cortical maps. *Neuron* **56**, 384–398 (2007).
69. Gigerenzer, G. & Hug, K. Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition* **43**, 127–171 (1992).
70. Karmiloff-Smith, A. *Beyond Modularity: A Developmental Perspective on Cognitive Science* (MIT press, 1995).
71. Sternberg, R. J. Domain-generality versus domain-specificity: the life and impending death of a false dichotomy. *Merrill-Palmer Quarterly* **35**, 115–130 (1989).
72. Insel, T. R. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am. J. Psychiatry* **171**, 395–397 (2014).
73. Insel, T. *et al.* Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
74. Rosch, E. & Lloyd, B. L. *Cognition and Categorization* (Lawrence Erlbaum Associates, 1978).
75. Barulli, D. & Stern, Y. Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends Cogn. Sci.* **17**, 502–509 (2013).
76. Kolb, B. *et al.* Experience and the developing prefrontal cortex. *Proc. Natl Acad. Sci. USA* **109** (Suppl. 2), 17186–17193 (2012).
77. Mueller, S. *et al.* Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586–595 (2013).
78. Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D. & Damasio, A. R. A neural basis for lexical retrieval. *Nature* **380**, 499–505 (1996).
79. Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. Neural correlates of category-specific knowledge. *Nature* **379**, 649–652 (1996).
80. Carrington, S. J. & Bailey, A. J. Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum. Brain Mapp.* **30**, 2313–2335 (2009).
81. Spunt, R. P. & Lieberman, M. D. The busy social brain: evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychol. Sci.* **24**, 80–86 (2013).
82. Spunt, R. P. & Lieberman, M. D. in *Dual-Process Theories of the Social Mind* (eds Sherman, J. W., Gawronski, B. & Trope, Y.) 279–299 (Guilford Press, 2014).
83. Clithero, J. A. & Rangel, A. Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
84. Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M. & Kanwisher, N. Domain specificity in visual cortex. *Cereb. Cortex* **16**, 1453–1461 (2006).
85. Saxe, R. & Kanwisher, N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835–1842 (2003).
86. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.* **2**, 568–573 (1999).
87. Tarr, M. J. & Gauthier, I. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.* **3**, 764–769 (2000).
88. Bilalic, M., Grottenthaler, T., Nägele, T. & Lindig, T. The faces in radiological images: fusiform face area supports radiological expertise. *Cereb. Cortex* **26**, 1004–1014 (2016).
89. Webster, M. A., Kaping, D., Mizokami, Y. & Duhamel, P. Adaptation to natural facial categories. *Nature* **428**, 557–561 (2004).
90. Heyes, C. Submentalizing: i am not really reading your mind. *Perspect. Psychol. Sci.* **9**, 131–143 (2014).
91. Kanwisher, N. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl Acad. Sci. USA* **107**, 11163–11170 (2010).
92. Duchaine, B. C., Dingle, K., Butterworth, E. & Nakayama, K. Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron* **43**, 469–473 (2004).