

What neuroscience can reveal about cognition and its origins

Skerry, A.E. and Saxe, R.
3.30.2015

Cognitive scientists have long been concerned with the origins of human knowledge. Over the last several decades, researchers have used the behavioral tools of developmental psychology to characterize the infant mind and chart the course of developmental change. Nonetheless, fundamental questions concerning the origins of human cognition remain open. In this chapter, we argue that the methodological and conceptual tools of neuroscience can provide unique insights, complimenting what has been learned from behavioral methods alone. We begin by highlighting several questions at stake in research on cognitive development: these questions concern 1) the extent to which human cognition depends on domain-specific vs. domain-general mechanisms 2) the precise form and content of mature cognitive representations, 3) the interplay between innate structure and learning in the emergence of these representations, and 4) possible continuities and discontinuities over development. We use two case studies—face perception and theory of mind—to review the challenges associated with each of these questions, and discuss how neuroscience can be brought to bear on them. For each case study, we argue that the tools of neuroscience have already made fruitful contributions, and sketch out ways that existing approaches could yield further progress in these domains.

INTRODUCTION

Over the last several decades, developmental psychologists have charted the development of human cognitive abilities with extraordinary breadth and precision (1). This research has revealed rich representational resources available remarkably early in life (2), as well as cases of surprisingly prolonged, piece-meal development (3). Still, despite this exciting progress, many of the most fundamental questions concerning the origins of human cognition remain unanswered.

Here, we propose that theories and methods of neuroscience could contribute to this endeavor, focusing on four classes of outstanding questions. First, there is a long-standing debate in cognitive science regarding human cognitive architecture (4,5): does the mind consist of a few general-purpose mechanisms carrying out diverse functions, or a larger number of functionally-specialized mechanisms tailored to computational demands of specific domains? For any given cognitive phenomenon, what are the contributions of domain-specific and general-purpose systems? Second, we must understand how the assumptions of the domain are represented, and how those representations support domain-relevant operations. Characterizing the form and content of mature cognitive representations is a prerequisite for understanding the developmental processes that yield them.

This brings us to a third question: what is the initial state of the human mind (i.e. what is specified genetically?) and what are the mechanisms that yield mature cognition from these initial resources? Mechanisms of change could include both experience-independent maturational mechanisms, as well as learning mechanisms that yield qualitatively new representational systems (1,6). For any domain, we want to understand the interplay between innate structure and learning in the emergence of adult-like representations. A final, related question concerns possible continuities and discontinuities in cognitive mechanisms over development. Does behavioral change correspond to the emergence of entirely new mechanisms, or on the tuning and refining of existing systems?

For each of these issues, we argue that the vocabulary, tools, and phenomena of neuroscience provide insights that complement what has been learned from behavioral methods alone. Given that

cognitive science ultimately aims to understand the human mind at multiple levels (e.g. computational, algorithmic, implementational; 7), characterizing the neural basis of human cognition is an important enterprise in its own right. However, in the present chapter, we make a stronger claim: that neuroscience can inform the level of analysis typically of interest to researchers in cognitive psychology and cognitive development (e.g. 1). While Marr argued that different levels of analysis could be pursued independently, a premise of modern cognitive neuroscience is that these levels inform and constrain one another, and that theories and phenomena from cognitive psychology can thus guide understanding of neural function. We propose that this relationship is partially symmetric, and that the study of cognitive development can benefit from theories and findings of neuroscience. We advance this proposal in the context of two case studies: face perception and theory of mind.

We use three general criteria as evidence for a “contribution” to the study of human cognitive development. First, neuroscience methods could provide evidence that converges with results of behavioral studies (e.g. looking time), possibly ruling out alternative explanations of the behavior. Second, neuroscience could enable comparison with relevant populations (e.g. different species, developmental history), allowing us to “look under the hood” in populations otherwise difficult to study effectively. Finally, and most excitingly, neuroscience could provide evidence that differs qualitatively from what we learn from behavior. In particular, we argue that neuroscience can provide a unique window into intermediate stages in a hierarchy or processing stream, allowing access to representations that are encapsulated from awareness or even behavior. This is itself a deep contribution to cognitive science, as it provides constraints on the algorithms by which information is transformed during processing and inference. Moreover, these insights can provide hypotheses and tools for charting the emergence of different levels of representation over human development, and can yield novel predictions for behavioral studies. In some cases, neuroscience has already made such contributions; in others, the data are not yet available and we focus on potential contributions. Our goal is not to provide a comprehensive review of developmental cognitive neuroscience, but instead to sketch a roadmap for research integrating

neuroscience, cognitive psychology, and developmental science to understand the human mind and brain.

FACE PERCEPTION

For a social species, faces are an essential source of information. A conspecific's face signals identity (8), attention (9), and emotion (10,11), and we can recognize hundreds of distinct faces, even under highly variable conditions (e.g. lighting, viewpoint, position: (12). What mechanisms subserve this ability, and how do they emerge in human development?

Domain-specificity.

Behaviorally, face processing exhibits signatures of a specialized perceptual mechanism (13–15). Nevertheless, cognitive scientists have long debated whether face-processing depends on mechanisms specialized for face perception, or on more general-purpose visual systems (16–19). Research in cognitive neuroscience has contributed to this debate by revealing large patches of the human cortical surface that are selectively involved in face perception. A region of the mid-fusiform gyrus (termed fusiform face area, FFA), for example, exhibits a reliably stronger response to faces compared to non-face objects (20,21). The face response is robust across stimulus sets and presentation parameters (22–24), responding more to identical sensory input when perceived as a face versus not (25). The FFA's response to faces is twice as strong as the response to any non-preferred stimulus (see (5,22) for review) and selectivity is observed even in the response of individual neurons (26,27).

Moreover, the response in the FFA is tightly linked to face-processing behaviors: it converges with well-established signatures of face perception from behavioral psychophysics (28–30), and is related to face detection and recognition performance on a trial-by-trial basis (31). Damage to the FFA is associated with selective deficits in face memory and recognition (32,33) and variability in the FFA response predicts individual differences in face recognition (34,35). While the FFA has been studied most extensively, many of these characteristics are shared by other face-selective regions, including the occipital face area (OFA) and a region of the superior temporal sulcus (STS) selective for dynamic faces (36,37). Together, an extensive body of research suggests that this set of cortical regions is responsible for

solving core computational challenges of detecting and recognizing conspecific faces.

Face representations and computations.

Beyond providing evidence for domain-specificity, this selectivity opens the door for investigating the internal workings of this cortical system. Neuroscientists are now well positioned to test how populations of neurons within face-selective regions perform the computations required for face processing (e.g. transforming retinal input into invariant identity representations that generalize across viewpoint and expression; (38). A precise characterization of these representations is a fundamental goal in its own right, and an essential step in charting the development of these mechanisms. Having identified a face-selective network, neuroscientists can now probe its intermediate stages, allowing access to representations difficult to study behaviorally.

Here, the most exciting work has come from monkey neurophysiology. Like humans, macaque monkeys exhibit selective cortical responses to faces (27), forming a set of six interconnected regions (39) that are causally involved in facial discrimination (40). Although the precise extent of homology between macaque and human face networks remains poorly understood (41,42), there appear to be deep parallels in functional organization across these two species (43,44)¹. Thus, the diverse tools of systems neuroscience can be used to characterize the representations in different face-selective regions. By characterizing the response profiles of single neurons (e.g. what stimulus parameters drive neuronal firing, and what is the tuning function like?), and the information that can be read out (“decoded”) from activity across a given population of neurons, this research can provide constraints on the algorithms that solve facial recognition in the primate brain. This, in turn, can provide a roadmap for understanding face perception in the human brain. An elegant line of work using these techniques has begun to piece together the series of representations that support face detection and face recognition.

¹ While some of speculated about regions in the macaque ventral stream that may be analogous to human FFA (43,44), it remains unclear whether there is a direct region-to-region mapping between the macaque and human face network (41,42).

The most posterior region of the macaque face network (posterior lateral: PL) is hypothesized to play a role in the initial detection of faces. A bottom-up investigation of the features that drive this region revealed a surprising profile: neurons in PL respond maximally to the presence of an eye surrounded by a face boundary in the top contralateral visual field (45). Many have proposed that face detection relies on a whole-face template, rather than combining simpler feature detectors (46–48). However, neural responses in PL were additive (cells responded to a face boundary and eye in isolation) and did not depend on face-appropriate configuration of the eye within the contour. Thus, the earliest stages of face detection may depend on summation over relatively simple, retinotopically localized eye-like features and contours, rather than conjunctions or whole-face templates (45).

More anterior regions of the macaque face network (middle lateral: ML, middle fundus: MF) appear to also play a role in face detection, but with a different profile. In one study (49), monkeys viewed artificial images that varied the local contrast relations between face parts (e.g. left eye versus nose). Neurons in regions ML/MF were strongly modulated by local signed-contrast relations (which patch is brighter and darker), with a high proportion of neurons coding contrast relations in the eye region. Interestingly, neuronal contrast preferences corresponded to the specific contrast differences that are computationally useful for detecting faces under realistic illumination conditions (50,51). Behavioral sensitivity to contrast polarity in faces (52,53)² may therefore stem from this intermediate stage of cortical face processing where neurons are tuned to local contrast regularities that pick out faces in natural scenes.

PL and ML/MF are the most posterior regions of the macaque face network. Recent research suggests that the six macaque face regions are organized into a posterior to anterior hierarchy of increasingly abstract representations of facial identity. For example, in ML/MF, neurons are tuned not only to local contrast (relevant for face detection), but also to face geometry (relevant for face identity). In one study (55), monkeys were presented with artificial face stimuli that parametrically varied along 19 dimensions (e.g. iris size, inter-eye distance). Neurons in ML/MF were selective for individual face parts,

² often attributed to a coarse, subcortical face-processing mechanism (52,54)

and tuned to the geometries of these face parts, with a particularly high frequency of tuning to facial layout parameters (e.g. face-aspect ratio) and eye geometry (e.g. iris size). Interestingly, neural responses to decomposed versions of these face stimuli exhibited significant second-order interactions between different face parts, indicating sensitivity to higher-level conjunctions. However, the neuronal responses did not require the context of a complete, upright face. This partial context specificity suggests that in ML/MF (unlike PL) facial parameters are not represented as isolated low-level features, but are integrated in a partially configural representation of a face (55).

In another study, monkeys viewed images of 25 individuals in eight different head orientations (56). In regions ML/MF, the neural patterns were strongly organized around head orientation. Images of the same orientation were most similar in their neural responses and only a small proportion of cells (19%) showed significant modulation by identity. Thus, although neurons in ML/MF code identity-relevant face parameters, the initial representation of that feature space may be relatively intolerant to transformations like head orientation.

In contrast, the most anterior region (anterior medial: AM) contained a majority (73%) of identity-tuned neurons, and these responses were highly tolerant to changes in viewpoint. Interestingly, a region that falls between these two patches, anterior lateral (AL), seems to form a mid-level representation that exhibits partial viewpoint invariance (56). Specifically, the pattern of response to same identity is most similar for mirror-symmetric views (e.g. right and left profile), but not for other changes in orientation. These results suggest that neurons in AL may reflect an intermediate representation pooling over mirror-symmetric views of a face, a discovery that may help clarify the algorithms that support invariant face recognition (38).

These data, combined with other findings of increased invariance across inferior temporal cortex (e.g. to position: (57)), lend strong support for a hierarchical arrangement of face-selective regions. Moreover, these results confirm that computations within the face-processing network are spatially segregated in distinct regions, and provide an initial blueprint for models of face recognition inspired by

this cortical hierarchy. In sum, neuroscience has compellingly demonstrated the existence of a domain-specific system, and begun to make exciting progress in piecing together the representational stages that contribute to face-relevant computations. Together, the research characterizing these regions' functional specificity and internal feature spaces provides the foundation for asking two related questions about the development of human face processing.

Developmental Trajectory.

How do human children develop sophisticated face-processing abilities? We argue that some of the deepest questions regarding the development of human face perception can be informed by the data, methods, and models of neuroscience. For example, is there any innate basis for human face perception and if so what are the format and content of these innate representations (question #3 above)? That is, which aspects of mature face-processing computational machinery are specified genetically and which aspects are derived from visual experience with faces? A related question (question #4) concerns continuity in the mechanisms that support face processing over the life span³. In most cases, the data required to answer these questions are not yet available, but we provide a sketch of a research program that could make progress.

Given the centrality of faces to successful social functioning, it is unsurprising that the primate visual system has evolved innate mechanisms for attending to and processing faces (5,58,59). Behavioral evidence shows that key face processing signatures are present remarkably early in life, with studies of infants replicating many of the key phenomena from adult psychophysics of face processing (60–63). Face sensitivity is not dependent on experience with faces, as preferences for faces over comparably structured non-face objects are present even in neonatal infants (46,52,64) and in animals raised with no visual experience with faces (65,66).

Still, the scope and extent of innate structure remain unclear. What is the precise computational

³ This question is related to, but distinct from the question of innate content. There could be relatively sophisticated innate abilities supported by a developmentally discontinuous system, or continuous systems that undergo substantial developmental change and refinement.

role of the innate face representations, and how do these relate to mature face-processing mechanisms? At one end of the spectrum, some have proposed that the cortical face-processing network is a genetic adaptation to complex social-communicative demands faced in recent primate evolution (67). An extreme interpretation of this hypothesis would be that the full structure of the domain-specific cortical mechanism is specified genetically, supporting all face-relevant computations (e.g. individuation, gaze detection, expression discrimination) prior to visual experience with faces (59,68). Alternatively, early face preferences could be supported by relatively minimal innate structure that precedes the cortical hierarchy described above (69). For example, neonates might possess only a crude perceptual template that biases attention towards faces, with the rest of the specialized system built from domain-general learning mechanisms operating over this preferentially attended input (46,70). Of course, there are a variety of intermediate possibilities between these extremes.

Behavioral evidence for early face processing is insufficient to test these two hypotheses. While many studies of face-processing have been conducted with preverbal infants, faces are a frequent if not predominant visual input over the early months (71,72), and it is known that face-processing abilities undergo experience-dependent change in the first year (63). Thus, studies with neonatal or face-deprived organisms are most relevant for questions about innate structure. However, most studies with neonates have focused on visual face preferences (e.g. (52), elucidating the features that guide infants' attentional bias for faces (e.g. non-scrambled, upright, appropriate contrast). Evidence for other face relevant computations is much more limited. Although studies have demonstrated successful recognition of individual faces (73) though see (74), none of these studies include a comparison to non-face objects, and therefore leave open the question of whether neonatal individuation relies on mechanisms that are face-specific⁴. Other proposed signatures of continuity are similarly ambiguous. For example, evidence that visual orientation to faces in early infancy predicts later individuation performance (75), or that face-

⁴ Moreover, a comparison with non-face objects would be difficult to interpret: if infants individuate faces but not non-face objects, might that be because faces are simply more interesting (given the attentional preference) or more visually discriminable?

processing abilities and disorders are highly heritable (76–78), could be explained either in terms of an underlying mechanism that is continuous across development, or an initial attentional bias that gates the development of a new, experience-dependent mechanism. We propose that neuroscience provides a rich alternative approach to disambiguate these hypotheses.

First, neuroscience can provide direct signatures for testing whether the same neural mechanism supports early and later abilities. In addition to the occipitotemporal face network described above, regions such as the superior colliculus, pulvinar, and amygdala have been argued to support a face detection system that is fast (79), coarse (80) and unconscious (81,82), perhaps homologous to subcortical mechanisms for orienting towards prey, predators, and conspecifics across a range of vertebrate taxa (83). Given that adult face-processing is supported by both subcortical and cortical systems, do infant face preferences depend on a genetically specified cortical specialization for face processing, or are can they be fully explained by more primitive subcortical face-processing mechanisms (69,84)? If the latter, what is the precise role of early visual experience (guided by the innate subcortical system) in the construction of new domain-specific mechanisms?

Given that these two mechanisms have well-characterized neural markers, the recruitment of different brain regions can serve as signatures for testing which face-processing mechanisms are operational early in development. A line of studies attempted to tackle this question by using fMRI to measure neural response to faces in children: while initial reports suggested prolonged development in the FFA (85,86), rigorously controlling for confounding variables like motion and temporal signal-to-noise reveals no reliable differences between children and adults (Dilks, personal communication; (87). Still, these children have years of exposure to human faces; what neural mechanisms support face-processing in infancy? Thus far, the dominant tool for investigating the neural basis of early face processing has been electroencephalogram (EEG). Although this method has poor spatial resolution, its high temporal resolution could be useful if cortical face regions and subcortical systems were known to differ in their temporal profiles. However, while many have associated the canonical N170 with cortical mechanisms

(and shorter latency components with subcortical processing (79)), the source of the N170 has yet to be conclusively determined (88–90). Moreover, some have suggested that seemingly equivalent components may have a different source in adults and children (59). Empirically, claims of face-specific EEG signals in infancy and childhood have also been inconsistent across studies. While some studies report an “infant N170”, this face-specific component occurs with ~100ms greater latency, and is more medially distributed compared to the adult N170. Furthermore, key properties of the N170 are absent at early ages: whereas a late positive component (P400) shows sensitivity to inversion in young infants, an inversion effect for the N170 does not emerge until 12 months of age (91,92), and other studies report gradual change in the N170 well into childhood (93). Methods such as fMRI and Near Infra-Red Spectroscopy (NIRS) provide promising alternatives for measuring neural responses to faces in infancy, as they have the spatial resolution to directly localize regions involved in early face processing.

Still, infants are heavily exposed to conspecific faces from birth, and face-processing mechanisms undergo substantial change during the first year (63). Thus, any imaging methods would need to be brought to very young, or even neonatal, infants in order to conclusively identify innate cortical mechanisms. A complimentary approach is to measure face responses in developmentally mature populations that lack visual experience with faces. For example, neuroscientists could measure neural responses in monkeys following selective deprivation for faces (e.g. (66)). If cortical face-selectivity is observed following controlled-rearing, this would provide strong evidence against the necessity of visual face experience in the emergence of the cortical face network (see also (94)).

So far, we have focused on the question of whether infants’ face processing abilities are supported, in part, by the cortical mechanism associated with adult face processing. However, there are a variety of more nuanced questions that neuroscience could help address. Specifically, neuroscience may be useful for characterizing the structure and content of innate face representations, and the mechanisms of developmental change. The electrophysiological results reviewed above offer tools for directly measuring neural feature spaces in experience-deprived monkeys, and hypotheses to guide behavioral studies with

human infants. In particular, these results make novel and nuanced predictions about intermediate stages of face-processing, which could motivate new behavioral experiments on the features that affect face processing in human neonates.

For example, neurons in PL and ML/MF are driven by specific perceptual parameters (eye in contour and local contrast respectively) that are diagnostic of faces in natural scenes. Do these response profiles emerge as a result of experience with statistical regularities of faces or are they part of an innate schema for face detection? This question could be addressed by directly measuring neural responses in face-selective patches of monkeys reared without face exposure, or via behavioral experiments with human infants. For example, in PL, a bottom-up approach revealed features of face representations that are neither simple visual parameters nor holistic perceptions—do these same stimulus properties drive infant attentional preferences? The research on facial contrast relations provides a particularly strong example of the way in which neuroscience can motivate and constrain developmental investigations. Whereas initial behavioral research with human infants focused on the simple distinction between forward and inverse contrast polarity (95), Ohayon and colleagues (49) demonstrated that the contrast relations that drive ML/MF neurons are specifically those relevant for picking out faces in natural scenes (e.g. (51)). A recent NIRS study shows that similar contrast sensitivities (particularly intact contrast relations between the sclera and iris (53)) are reflected in neural responses in infant temporal cortex (96).

Finally, the methods of neuroscience allow us to examine not only the content of innate representations, but also change/continuity in these representations over development. For example, key signature of neural face representations is the emergence of viewpoint-invariant identity representations in anterior face regions (56)—does viewpoint-invariance depend on visual experience with faces (see (97)? Neonates can reportedly recognize faces across certain viewpoints (73), but it is unknown whether this ability depends on viewpoint-invariant face representations of the kind documented in AM. Does developmental change in perceptual invariances reflect the emergence of distinct representations (e.g. a new cortical region), or altered tuning within existing neural populations?

A related example comes from the literature on perceptual narrowing. Over the first 9 months of life, infants' recognition abilities become restricted to faces of the race and species to which infants are exposed (98,99). Should we think of this behavioral change in terms of tuning of underlying face feature spaces within a developmentally continuous system? Given that individual neurons in ML/MF respond parametrically to specific face features (55), a tempting interpretation of perceptual narrowing would be that it alters the neural feature space in such regions. For example, there could be an innate feature space for face individuation that is rich but broadly tuned, and perceptual narrowing might reflect the pruning of the face-specific features (or ranges of values on those features) to prioritize coding of feature values that are diagnostic of identity in the infant's face environment. An alternative possibility is that the innate feature space is very minimal and infants rely on domain-general representations to discriminate faces. On this account, perceptual narrowing might reflect a shift to a new set of representations altogether (a developmental discontinuity), as neurons develop preferential responses to specific geometric parameters of faces. In principle, the methods of primate neurophysiology offer an exciting avenue for distinguishing these possibilities. By combining neuroscience with controlled-rearing methods from comparative and developmental psychology (e.g. (66)), we may be able to understand not only the nature of such representational changes, but also the maturational and experiential influences that drive them.

In sum, neuroscience has deepened our understanding of face processing and its emergence over human development in several ways. First, it has demonstrated the existence of domain-specific mechanisms for processing faces, backed up by rigorous tests of face-selectivity. Second, it has provided a more precise description of different stages and levels of representation involved in solving the key computational challenges of face detection and recognition. Third, it has validated the existence of distinct cortical and subcortical face-processing systems, and provided reliable signatures to pinpoint the mechanisms in operation at different ages. And finally, it has provided tools with which we can directly measure the feature spaces and representational distinctions present in innate face processing machinery, and characterize change in those feature spaces over development.

THEORY OF MIND

In this section, we explore the role of neuroscience in the study of a second social function: the ability to reason about the minds of others using a so-called 'Theory of Mind' (ToM; (100,101). To explain and predict the behavior of other people, we rely on a coherent set of intuitive principles about how other minds work. This intuitive theory specifies abstract causal relations between different mental states (e.g. how desires, expectations, and beliefs determine emotions), between mental states and events or circumstances in the world (e.g. how a person's perceptual experiences shape her beliefs), and between mental states and overt behavior (e.g. that people act so as to bring about their desires and intentions efficiently). Equipped with these "theory-like" assumptions, we can go beyond observed behaviors to recover hidden internal states, and recruit inferred mental states for a range of relevant inferences.

Unlike a domain of high-level vision such as face perception, our knowledge of other minds thus has rich causal structure, expressed over abstract concepts, and is largely independent of stimulus properties. The methods of neuroscience described above have primarily been developed for characterizing perceptual systems. Can these methods also make progress in the study of abstract, causal domains? Theory of mind provides a useful case study for exploring the contributions of current neuroscientific methods to the study of high-level cognitive processes, as well as the limits of extant approaches.

Domain-specificity.

On its face, ToM has properties in common with many domains of abstract knowledge. For example, we might rely on an intuitive theory of the weather to predict that it will rain (e.g. based on the cloudy sky), and to make inferences about where the storm will move (e.g. based on the wind). Is the brain equipped with distinct selective mechanisms for each domain of abstract knowledge, or are diverse domains of reasoning supported by general-purpose mechanisms that support prediction, inference, and explanation regardless of content? While a domain-general mechanism might seem more plausible, a number of lines of evidence suggest that mental state reasoning is "special". For example, theory of mind

exhibits a unique and systematic developmental trajectory from infancy through late childhood (102), and the development of this ability is selectively impaired in individuals with Autism Spectrum Disorders (103).

Again, findings from neuroscience provide strong convergent evidence in favor of a domain-specific mechanism. A striking discovery in cognitive neuroscience has been a set of cortical regions—medial prefrontal cortex (MPFC), precuneus (PC), bilateral temporal-parietal junction (TPJ), and bilateral anterior temporal cortex (ATL)—that respond selectively to reasoning about other’s mental states (104,105). This selective response is robust and replicable: the ToM regions can be identified in 80-90% of individual subjects in just 10 minutes of scan time, and can be identified with a wide range of stimuli and tasks (e.g. (106,107); see (108) for review). The region most selective for mental state content appears to be the rTPJ; this region shows a higher response when reasoning about a person’s mental states compared to the same person’s bodily states or physical appearance (109), does not vary with the logical and executive functioning demands of the task (104,110), and is distinct from nearby regions involved in representing bodies, actions, and goals (111–113).

The surprisingly selective neural substrate for theory of mind raises further questions concerning the nature and origins of domain-specificity. What drives the development of functionally-specific circuitry, and why do some processes get their own large scale, functionally specialized cortical real estate while others do not (114)? Unlike face processing, reasoning about others’ mental states (particularly representational states like false beliefs) may not be phylogenetically ancient, nor developmentally primitive. By traditional measures, it is not until age 4 that children understand that a person’s behavior is guided by her beliefs ((102); though see below), and the ability to reason about others’ false beliefs appears to be either uniquely human (115,116) or present in few other species (117). Yet, like perceiving a face, reasoning about what someone is thinking or feeling appears to rely on a reliable, highly specialized neural substrate. While we know little about what drives emergence of selectivity for theory of mind, it does appear to be behaviorally relevant: emerging selectivity in the rTPJ is predictive of behavior

on a battery of challenging theory of mind problems (118).

ToM representations and computations.

As with face processing, identifying a network of selective regions is only the first step in figuring out what these regions do. More recent research has investigated the specific aspects of social cognition that drive the response in particular regions: for example, the rTPJ shows the strongest response for propositional mental states like beliefs, while the MPFC responds strongly to information about traits, emotions, and preferences (109,119,120). However, these studies focus on the magnitude of response across entire regions. As we observed for face perception, many of the most exciting discoveries can come from fine-grained investigations of the response of individual neurons, or patterns of activity across populations of neurons. For a human-specific ability like ToM, the invasive tools of neurophysiology are not feasible. Luckily, a growing literature suggests the same basic approach can be extended to functional MRI, even though it lacks the spatial resolution of single-unit recordings. Methods such as multi-voxel pattern analysis (MVPA) allow us to characterize the information contained in distributed patterns of activity within a cortical region (121,122), meaning we can potentially use neuroscience to characterize representations even in domains like theory of mind.

A new line of studies using this approach has identified several abstract dimensions of mental states that can be decoded from the patterns of activity across the rTPJ. Specifically, neural patterns in the rTPJ distinguish accidental vs. intentional harmful acts (123), the perceptual modality through which others' beliefs are formed (visual vs. auditory; (124), and the epistemic validity (good vs. poor evidence) of different beliefs (125); the discriminability of these neural patterns predicts individual differences in behavior (123). Together, this growing body of research shows that it is possible to probe fine-grained features represented in different theory of mind regions.

This research has focused on epistemic properties of beliefs, but what about the affective or emotional properties of others' mental states? A number of studies implicate MPFC in emotion processing (120), and recent findings suggest that distributed patterns of activity in the MPFC contain

information about distinct emotions when presented in facial expressions, bodily motions, or tones of voice (126). More recently, we found that patterns of activity in MPFC reflect an abstract representation of emotion that can be elicited even by a schematic depiction of the person's situation (e.g. being excluded from a social group) even without any overt expression (127). In contrast, patterns of activity in the FFA could distinguish between positive and negative emotion only when presented in facial expressions. Finally, a region in posterior temporal cortex (STC), appears to contain a representation of valence that generalizes across modalities (i.e. faces, bodies and voices in (126) but does not distinguish trials in which positive versus negative affect are conveyed via the situation (127). Thus, reminiscent of the intermediate face representations observed in AL (56), STC may contain a representation of others emotions that is only partially invariant to properties of the stimulus.

In sum, this emerging body of research suggests that, as with face perception, the tools of neuroscience allow us to make progress on the internal workings of neural systems selective for ToM. We can identify the various discriminations supported by different regions, and start to understand the way those discriminations are constructed within a hierarchy of increasingly abstract, invariant representations.

Developmental Trajectory.

Can a deepened understanding of the neural processing stream for theory of mind inform our understanding of theory of mind development? Despite being one of the most extensively studied domains, the development of theory of mind is, in many ways, still a mystery. For decades, behavioral data suggested a dramatic change in the ability to reason about other's beliefs occurring around age four. In the classic Sally-Anne paradigm, for example, a child is introduced to a character's belief about the location of her toy (e.g. Sally puts her toy under her bed) followed by an event that renders this belief false (e.g. while Sally is outside, Anne moves Sally's toy to the closet). When asked where Sally will look for her toy, children under the age of four point to the true location of the toy, whereas older children recognize that Sally will act on the basis of her outdated belief (128). This phenomenon replicates across a variety of paradigms and tasks, and has been taken as evidence for a key conceptual change in

understanding other minds (102). However, more recent studies using measures of spontaneous looking-time have revealed expectations about others' beliefs within the first year of life (129). These new data raise deep puzzles and challenges, which neuroscientific tools may help address (130).

Why do three-year-old children apparently fail to take others' beliefs into account in their explicit predictions and explanations of others' actions? One possibility is that around four years, children develop a new concept of other's beliefs (101,102), integrated with their previously existing theory of other's desires, intentions, and actions. On this view, the difference between three and five-year-olds reflects a reorganization of concepts within children's theory of mind, and therefore plausibly the emergence of distinct patterns of activity within existing brain regions, or the emergence of new regions within the ToM network (e.g. the emergence of the right TPJ).

An alternative account is that children's concepts of others' beliefs may be stable over this period (129,131), with improved performance on typical tests of theory of mind reflecting the maturation of domain-general resources that unmask existing knowledge (e.g. children develop the inhibitory control required to suppress prepotent tendencies to point where the object actually is: (132,133). If so, performance on theory of mind tasks should be correlated with maturation of the brain systems that are involved in executive function and inhibitory control (including dorso-lateral prefrontal cortex (DLPFC) and the intraparietal sulcus (IPS) e.g. (134). This alternative view does not predict any association between task performance around age four years and activity in brain regions associated with theory of mind, like rTPJ and MPFC.

Although the critical tests have not been conducted, initial hints seem to favor the conceptual change hypothesis: in four-year-olds, performance on standard theory of mind tasks is specifically related to maturation of the rTPJ and MPFC (135). However, in this study, maturation was inferred from an EEG measure of coherence of alpha waves at rest, where increased coherence is thought to reflect biological maturation processes such as myelination. Thus, this measure cannot directly test whether representations in the rTPJ and MPFC are changing. In older children aged five to twelve years, performance on an

expanded battery of theory of mind tasks (including moral evaluations of accidents based on false beliefs) is correlated with change in the rTPJ and MPFC, and not with change in DLPFC or IPS (118,136), and the rTPJ in particular showed increased selectivity in its response to beliefs, as opposed to other social information. However, these studies measured only the magnitude of response in each region, and so provide limited access to the representations involved. Thus, it remains to future studies to test whether maturation in executive function brain regions, or altered representations in rTPJ and MPFC, are better predictors of changing performance on explicit tests of theory of mind.

A second set of questions concerns the abilities of younger children and even infants. Does infant sensitivity to others' beliefs depend on the same mechanism that adults recruit explicitly reasoning about others' beliefs, or are early theory of mind abilities supported by a distinct (possibly fast and automatic) mechanism, as some have argued for early face perception? If the latter, does this "implicit" ToM mechanism continue to operate in adulthood, in parallel with the cortical theory of mind regions involved in explicit mental state reasoning (137–139)?

In principle, neuroscientific approaches could be used to address each of these questions, though in practice there are methodological hurdles. For example, in behavioral research, the key signature of belief understanding is that infants form differential action expectations when the other person has a false, versus true, belief (e.g. (140,141)). Recent studies have aimed to identify neural mechanisms involved in spontaneous or "implicit" theory of mind in adults using modified versions of infant behavioral stimuli that rely on this contrast between true and false beliefs (142,143). However, in adults, reasoning about *both* false and true beliefs depends on (and elicits equivalent) activity in theory of mind brain regions (144). Thus, a challenge for studying both theory of mind in infants, and a putative "implicit" theory of mind in adults, will be to design an experimental condition that elicits spontaneous theory of mind, and a control condition which does not, but which is matched for other low-level properties of the stimulus. Newer experimental designs that focus on testing representational distinctions within theory of mind regions (see below), rather than comparing conditions that do and do not elicit theory of mind, may

provide a way out of this conundrum.

As we further understand the internal workings of the theory of mind network (i.e. what different regions compute and represent), we can ask not only whether there is a region of cortex responsive to mental states continuously over development, but also more precise questions regarding which representations and computations are present innately, and which are constructed from specific kinds of experiences. For example, an exciting line of research has begun to investigate one type of experience that might affect the development of the ToM network: first-person visual experience. A dominant theory of human social cognition proposes that mental state reasoning involves “embodied simulations” that are dependent on first-person sensory experiences (145,146). On this proposal, mechanisms for reasoning about others visual percepts and knowledge should depend on one’s own experiences with seeing (147,148). Contrary to this hypothesis, congenitally blind individuals exhibit neural activity comparable to the sighted participants when listening to stories describing mental states (149). This was true even for mental states were formed on the basis visual percepts, and, as in sighted subjects, the rTPJ of congenitally blind individuals contains information about the modality from which a belief is formed (visual vs. auditory; (124). The striking invariance of ToM mechanisms to dramatic changes in developmental experience suggests that pressures independent of visual experience drive the emergence of this domain-specific cortical mechanism and its internal features. Are theory of mind mechanisms similarly invariant to other changes in developmental experience (e.g. language delays in congenitally deaf children: (150), and if not, which regions and representations are disrupted? Of course, research on the neural representation of theory of mind is at an early stage; we are only beginning to understand the internal computations of the mechanisms supporting ToM in adults. But as we build more nuanced hypotheses about different processing stages and corresponding neural features, this may yield methods and predictions for testing which representations are available in infancy, and for investigating the kinds of experience that yield change in those representations over development.

CONCLUSIONS

While the synergy of cognitive development and neuroscience holds tremendous promise, extant neuroscientific approaches are not without limitations. First, the research described here generally involved identifying “features” of different neural representations (e.g. the geometry of different face parts, the reliability of a belief). While this feature-based approach has been productive, it is unlikely that representations in all domains, particularly domains of high-level cognition such as theory of mind, can be reduced to operations over lists of associated features (151,152). Recently, computational cognitive scientists have formalized theory of mind and its development in terms of probabilistic inference over structured, generative knowledge (e.g. rational planning models: (153)). Thus, to capture the richly causal and compositional nature of the representations in domains like theory of mind, neuroscientists will need to move beyond a feature-based approach, incorporating the abstract, structured knowledge representations that have been fruitful in other areas of cognitive science (154). Merging the efforts of developmental science and neuroscience will raise methodological challenges as well. For example, many of the future directions proposed this chapter will depend either on improving methods for studying neural activity noninvasively in young infants (e.g. NIRS, infant fMRI), or on identifying appropriate animal models and combining developmental methods such as controlling rearing (e.g. (66) with the tools of neurophysiology (e.g. (55,56)).

Despite these challenges, the progress already made in the cases of face perception and theory of mind leave us optimistic about this approach. In this chapter, we have reviewed a number of recent discoveries from neuroscience that shed light on the mechanisms underlying both face perception and theory of mind. Moreover, we have sketched out future extensions of this work that may deepen our understanding of the development of social perception and cognition. Neuroscience lets us characterize different stages in processing, including intermediate stages encapsulated from behavior, and allows us to identify features that govern the representations at different stages. This in turn, can be used to further understand the development of these abilities.

REFERENCES

1. Carey S. *The Origin of Concepts*. Oxford University Press; 2009. 609 p.
2. Kinzler KD, Spelke ES. Core systems in human cognition. In: C. von Hofsten and K. Rosander, editor. *Progress in Brain Research* [Internet]. Elsevier; 2007 [cited 2014 Oct 14]. p. 257–64. Available from: <http://www.sciencedirect.com/science/article/pii/S007961230764014X>
3. Le Corre M, Carey S. One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*. 2007 Nov;105(2):395–438.
4. Fodor JA. *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press; 1983. 162 p.
5. Kanwisher N. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc Natl Acad Sci*. 2010 Jun 22;107(25):11163–70.
6. Carey S. *Conceptual Change in Childhood*. MIT Press; 1985. 226 p.
7. Marr D. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. W.H. Freeman; 1982. 397 p.
8. Bruce V, Young A. Understanding face recognition. *Br J Psychol Lond Engl* 1953. 1986 Aug;77 (Pt 3):305–27.
9. Senju A, Johnson MH. The eye contact effect: mechanisms and development. *Trends Cogn Sci*. 2009 Mar;13(3):127–34.
10. Darwin C. *The expression of the emotions in man and animals* [Internet]. London, John Murray; 1872 [cited 2014 Oct 14]. 410 p. Available from: <http://archive.org/details/expressionofemot1872darw>
11. Ekman P. Facial expression and emotion. *Am Psychol*. 1993;48(4):384–92.
12. Pinto N, Cox DD, DiCarlo JJ. Why is Real-World Visual Object Recognition Hard? *PLoS Comput Biol*. 2008 Jan 25;4(1):e27.
13. Moscovitch M, Winocur G, Behrmann M. What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. *J Cogn Neurosci*. 1997;9(5):555–604.
14. Tanaka JW, Sengco JA. Features and their configuration in face recognition. *Mem Cognit*. 1997 Sep;25(5):583–92.
15. Yin RK. Looking at upside-down faces. *J Exp Psychol*. 1969;81(1):141–5.
16. Diamond R, Carey S. Why faces are and are not special: an effect of expertise. *J Exp Psychol Gen*. 1986 Jun;115(2):107–17.
17. Farah MJ, Wilson KD, Drain M, Tanaka JN. What is “special” about face perception? *Psychol Rev*. 1998;105(3):482–98.

18. Gauthier I, Williams P, Tarr MJ, Tanaka J. Training “greeble” experts: a framework for studying expert object recognition processes. *Vision Res.* 1998 Aug;38(15-16):2401–28.
19. McKone E, Kanwisher N, Duchaine BC. Can generic expertise explain special processing for faces? *Trends Cogn Sci.* 2007 Jan;11(1):8–15.
20. Kanwisher N, McDermott J, Chun MM. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J Neurosci.* 1997 Jun 1;17(11):4302–11.
21. McCarthy G, Puce A, Gore JC, Allison T. Face-Specific Processing in the Human Fusiform Gyrus. *J Cogn Neurosci.* 1997 Oct 1;9(5):605–10.
22. Kanwisher N, Yovel G. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc B Biol Sci.* 2006 Dec 29;361(1476):2109–28.
23. O’Craven KM, Kanwisher N. Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *J Cogn Neurosci.* 2000 Nov;12(6):1013–23.
24. Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci.* 2008 Mar 18;105(11):4447–52.
25. Tong F, Nakayama K, Vaughan JT, Kanwisher N. Binocular Rivalry and Visual Awareness in Human Extrastriate Cortex. *Neuron.* 1998 Oct;21(4):753–9.
26. Parvizi J, Jacques C, Foster BL, Withoft N, Withoft N, Rangarajan V, et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci Off J Soc Neurosci.* 2012 Oct 24;32(43):14915–20.
27. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. A Cortical Region Consisting Entirely of Face-Selective Cells. *Science.* 2006 Feb 3;311(5761):670–4.
28. Schiltz C, Rossion B. Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage.* 2006 Sep;32(3):1385–94.
29. Yovel G, Kanwisher N. The Neural Basis of the Behavioral Face-Inversion Effect. *Curr Biol.* 2005 Dec 24;15(24):2256–62.
30. McKone E, Robbins R. Are Faces Special. In: Calder A, Rhodes G, Johnson M, Haxby J, editors. *Oxford Handbook of Face Perception.* Oup Oxford; 2011. p. 149–76.
31. Grill-Spector K, Knouf N, Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci.* 2004 May;7(5):555–62.
32. Riddoch MJ, Johnston RA, Bracewell RM, Boutsen L, Humphreys GW. Are faces special? A case of pure prosopagnosia. *Cogn Neuropsychol.* 2008 Feb;25(1):3–26.
33. 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 Aug 1;71(2):254–7.

34. Furl N, Garrido L, Dolan RJ, Driver J, Duchaine B. Fusiform Gyrus Face Selectivity Relates to Individual Differences in Facial Recognition Ability. *J Cogn Neurosci*. 2010 Jul 9;23(7):1723–40.
35. Huang L, Song Y, Li J, Zhen Z, Yang Z, Liu J. Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Front Hum Neurosci* [Internet]. 2014 Jul 2 [cited 2014 Oct 14];8. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4078911/>
36. Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*. 2011 Jun 15;56(4):2356–63.
37. Pitcher D, Walsh V, Duchaine B. The role of the occipital face area in the cortical face perception network. *Exp Brain Res*. 2011 Apr 1;209(4):481–93.
38. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron*. 2012 Feb 9;73(3):415–34.
39. Moeller S, Freiwald WA, Tsao DY. Patches with Links: A Unified System for Processing Faces in the Macaque Temporal Lobe. *Science*. 2008 Jun 6;320(5881):1355–9.
40. Afraz A, Boyden ES, DiCarlo JJ. Optogenetic and pharmacological suppression of face-selective neurons reveal their causal role in face discrimination behavior. *J Vis*. 2014 Aug 22;14(10):600–600.
41. Sereno MI, Tootell RBH. From monkeys to humans: what do we now know about brain homologies? *Curr Opin Neurobiol*. 2005 Apr;15(2):135–44.
42. Yovel G, Freiwald WA. Face recognition systems in monkey and human: are they the same thing? *F1000Prime Rep* [Internet]. 2013 Apr 2 [cited 2014 Aug 5];5. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3619156/>
43. Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci*. 2008 Dec 9;105(49):19514–9.
44. Rajimehr R, Young JC, Tootell RBH. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc Natl Acad Sci*. 2009 Feb 10;106(6):1995–2000.
45. Issa EB, DiCarlo JJ. Precedence of the eye region in neural processing of faces. *J Neurosci Off J Soc Neurosci*. 2012 Nov 21;32(47):16666–82.
46. Morton J, Johnson MH. CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychol Rev*. 1991;98(2):164–81.
47. Tanaka JW, Farah MJ. Parts and wholes in face recognition. *Q J Exp Psychol A*. 1993 May;46(2):225–45.
48. Tsao DY, Livingstone MS. Mechanisms of face perception. *Annu Rev Neurosci*. 2008;31:411–37.

49. Ohayon S, Freiwald WA, Tsao DY. What makes a cell face selective? The importance of contrast. *Neuron*. 2012 May 10;74(3):567–81.
50. Gilad S, Meng M, Sinha P. Role of ordinal contrast relationships in face encoding. *Proc Natl Acad Sci*. 2009 Mar 10;pnas.0812396106.
51. Sinha P. Qualitative Representations for Recognition. *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision [Internet]*. London, UK, UK: Springer-Verlag; 2002 [cited 2014 Oct 14]. p. 249–62. Available from: <http://dl.acm.org/citation.cfm?id=648248.751730>
52. Farroni T, Johnson MH, Menon E, Zulian L, Faraguna D, Csibra G. Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proc Natl Acad Sci U S A*. 2005 Nov 22;102(47):17245–50.
53. Otsuka Y, Motoyoshi I, Hill HC, Kobayashi M, Kanazawa S, Yamaguchi MK. Eye contrast polarity is critical for face recognition by infants. *J Exp Child Psychol*. 2013 Jul;115(3):598–606.
54. Stein T, Peelen MV, Sterzer P. Adults' Awareness of Faces Follows Newborns' Looking Preferences. *PLoS ONE*. 2011 Dec 21;6(12):e29361.
55. Freiwald WA, Tsao DY, Livingstone MS. A face feature space in the macaque temporal lobe. *Nat Neurosci*. 2009 Sep;12(9):1187–96.
56. Freiwald WA, Tsao DY. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*. 2010 Nov 5;330(6005):845–51.
57. Rust NC, DiCarlo JJ. Selectivity and Tolerance (“Invariance”) Both Increase as Visual Information Propagates from Cortical Area V4 to IT. *J Neurosci*. 2010 Sep 29;30(39):12978–95.
58. Sugita Y. Innate face processing. *Curr Opin Neurobiol*. 2009 Feb;19(1):39–44.
59. Mckone E, Crookes K, Kanwisher N. The cognitive and neural development of face recognition in humans. In M S Gazzaniga (Ed), *The cognitive neurosciences IV*. MIT Press; 2009.
60. Chien SH-L. No more top-heavy bias: Infants and adults prefer upright faces but not top-heavy geometric or face-like patterns. *J Vis*. 2011 May 19;11(6):13.
61. De Haan M, Johnson MH, Maurer D, Perrett DI. Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cogn Dev*. 2001 Apr;16(2):659–78.
62. Gliga T, Dehaene-Lambertz G. Structural Encoding of Body and Face in Human Infants and Adults. *J Cogn Neurosci*. 2005 Aug 1;17(8):1328–40.
63. Kelly DJ, Quinn PC, Slater AM, Lee K, Gibson A, Smith M, et al. Three-month-olds, but not newborns, prefer own-race faces. *Dev Sci*. 2005 Nov 1;8(6):F31–6.
64. Goren CC, Sarty M, Wu PY. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*. 1975 Oct;56(4):544–9.

65. Rosa Salva O, Farroni T, Regolin L, Vallortigara G, Johnson MH. The Evolution of Social Orienting: Evidence from Chicks (*Gallus gallus*) and Human Newborns. *PLoS ONE*. 2011 Apr 20;6(4):e18802.
66. Sugita Y. Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad Sci*. 2008 Jan 8;105(1):394–8.
67. Leopold DA, Rhodes G. A Comparative View of Face Perception. *J Comp Psychol Wash DC* 1983. 2010 Aug;124(3):233–51.
68. Farah MJ, Rabinowitz C, Quinn GE, Liu GT. Early commitment of neural substrates for face recognition. *Cogn Neuropsychol*. 2000 Feb 1;17(1):117–23.
69. Johnson MH. Subcortical face processing. *Nat Rev Neurosci*. 2005 Oct;6(10):766–74.
70. De Schonen S, Mathivet E. First come, first served: A scenario about the development of hemispheric specialization in face recognition during infancy. *Cah Psychol Cogn Psychol Cogn*. 1989;9(1):3–44.
71. Frank MC, Amso D, Johnson SP. Visual search and attention to faces during early infancy. *J Exp Child Psychol*. 2014 Feb;118:13–26.
72. Sugden NA, Mohamed-Ali MI, Moulson MC. I spy with my little eye: typical, daily exposure to faces documented from a first-person infant perspective. *Dev Psychobiol*. 2014 Feb;56(2):249–61.
73. Turati C, Bulf H, Simion F. Newborns' face recognition over changes in viewpoint. *Cognition*. 2008 Mar;106(3):1300–21.
74. Gliga T, Dehaene-Lambertz G. Development of a view-invariant representation of the human head. *Cognition*. 2007 Feb;102(2):261–88.
75. De Klerk CCJM, Gliga T, Charman T, Johnson MH, The BASIS team. Face engagement during infancy predicts later face recognition ability in younger siblings of children with autism. *Dev Sci*. 2014 Jul 1;17(4):596–611.
76. Duchaine B, Germine L, Nakayama K. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cogn Neuropsychol*. 2007 Jun;24(4):419–30.
77. Polk TA, Park J, Smith MR, Park DC. Nature versus Nurture in Ventral Visual Cortex: A Functional Magnetic Resonance Imaging Study of Twins. *J Neurosci*. 2007 Dec 19;27(51):13921–5.
78. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, et al. Human face recognition ability is specific and highly heritable. *Proc Natl Acad Sci*. 2010 Mar 16;107(11):5238–41.
79. Bailey AJ, Braeutigam S, Jousmäki V, Swithenby SJ. Abnormal activation of face processing systems at early and intermediate latency in individuals with autism spectrum disorder: a magnetoencephalographic study. *Eur J Neurosci*. 2005 May;21(9):2575–85.

80. Vuilleumier P, Armony JL, Driver J, Dolan RJ. Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat Neurosci*. 2003 Jun;6(6):624–31.
81. Pasley BN, Mayes LC, Schultz RT. Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron*. 2004 Apr 8;42(1):163–72.
82. Williams MA, Morris AP, McGlone F, Abbott DF, Mattingley JB. Amygdala Responses to Fearful and Happy Facial Expressions under Conditions of Binocular Suppression. *J Neurosci*. 2004 Mar 24;24(12):2898–904.
83. Sowards TV, Sowards MA. Innate visual object recognition in vertebrates: some proposed pathways and mechanisms. *Comp Biochem Physiol A Mol Integr Physiol*. 2002 Aug;132(4):861–91.
84. Morris JS, Buchel C, Dolan RJ. Parallel Neural Responses in Amygdala Subregions and Sensory Cortex during Implicit Fear Conditioning. *NeuroImage*. 2001 Jun;13(6):1044–52.
85. Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JDE, et al. Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat Neurosci*. 2007 Apr;10(4):512–22.
86. Scherf KS, Behrmann M, Humphreys K, Luna B. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev Sci*. 2007 Jul 1;10(4):F15–30.
87. McKone E, Crookes K, Jeffery L, Dilks DD. A critical review of the development of face recognition: Experience is less important than previously believed. *Cogn Neuropsychol*. 2012;29(1-2):174–212.
88. Bentin S, Allison T, Puce A, Perez E, McCarthy G. Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*. 1996 Nov;8(6):551–65.
89. Itier RJ, Taylor MJ. Source analysis of the N170 to faces and objects. *Neuroreport*. 2004 Jun 7;15(8):1261–5.
90. Nguyen VT, Cunnington R. The superior temporal sulcus and the N170 during face processing: single trial analysis of concurrent EEG-fMRI. *NeuroImage*. 2014 Feb 1;86:492–502.
91. De Haan M, Humphreys K, Johnson MH. Developing a brain specialized for face perception: A converging methods approach. *Dev Psychobiol*. 2002 Apr 1;40(3):200–12.
92. Halit H, de Haan M, Johnson MH. Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*. 2003 Jul;19(3):1180–93.
93. Taylor MJ, Batty M, Itier RJ. The faces of development: a review of early face processing over childhood. *J Cogn Neurosci*. 2004 Oct;16(8):1426–42.
94. Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A. Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron*. 2009 Aug 13;63(3):397–405.

95. Farroni T, Csibra G, Simion F, Johnson MH. Eye contact detection in humans from birth. *Proc Natl Acad Sci*. 2002 Jul 9;99(14):9602–5.
96. Ichikawa H, Otsuka Y, Kanazawa S, Yamaguchi MK, Kakigi R. Contrast reversal of the eyes impairs infants' face processing: a near-infrared spectroscopic study. *Neuropsychologia*. 2013 Nov;51(13):2556–61.
97. Li N, DiCarlo JJ. Unsupervised Natural Visual Experience Rapidly Reshapes Size Invariant Object Representation in Inferior Temporal Cortex. *Neuron*. 2010 Sep 23;67(6):1062–75.
98. Anzures G, Quinn PC, Pascalis O, Slater AM, Tanaka JW, Lee K. Developmental Origins of the Other-Race Effect. *Curr Dir Psychol Sci*. 2013 Jun 1;22(3):173–8.
99. Kelly DJ, Quinn PC, Slater AM, Lee K, Ge L, Pascalis O. The Other-Race Effect Develops During Infancy Evidence of Perceptual Narrowing. *Psychol Sci*. 2007 Dec 1;18(12):1084–9.
100. Gopnik A, Wellman HM. Why the Child's Theory of Mind Really Is a Theory. *Mind Lang*. 1992;7(1-2):145–71.
101. Perner J. *Understanding the Representational Mind*. MIT Press; 1991. 348 p.
102. Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev*. 2001 Jun;72(3):655–84.
103. Baron-Cohen S. Theory of mind and autism: A fifteen year review. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ, editors. *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed). New York, NY, US: Oxford University Press; 2000. p. 3–20.
104. Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. fMRI item analysis in a theory of mind task. *NeuroImage*. 2011 Mar 15;55(2):705–12.
105. Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage*. 2003 Aug;19(4):1835–42.
106. Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD. Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*. 2000;38(1):11–21.
107. Sommer M, Döhnell K, Sodian B, Meinhardt J, Thoermer C, Hajak G. Neural correlates of true and false belief reasoning. *NeuroImage*. 2007 Apr 15;35(3):1378–84.
108. Koster-Hale J, Saxe R. Functional Neuroimaging of Theory of Mind. In: Baron-Cohen S, Lombardo M, Tager-Flusberg H, editors. *Understanding Other Minds*. 3rd ed. 2013.
109. Saxe R, Powell LJ. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci*. 2006 Aug;17(8):692–9.
110. Saxe R, Schulz LE, Jiang YV. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc Neurosci*. 2006;1(3-4):284–98.

111. Downing PE, Jiang Y, Shuman M, Kanwisher N. A Cortical Area Selective for Visual Processing of the Human Body. *Science*. 2001 Sep 28;293(5539):2470–3.
112. Pelphrey KA, Morris JP. Brain Mechanisms for Interpreting the Actions of Others From Biological-Motion Cues. *Curr Dir Psychol Sci*. 2006 Jun;15(3):136–40.
113. Saxe R, Xiao D-K, Kovacs G, Perrett DI, Kanwisher N. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*. 2004;42(11):1435–46.
114. Downing PE, Chan AW-Y, Peelen MV, Dodds CM, Kanwisher N. Domain Specificity in Visual Cortex. *Cereb Cortex*. 2006 Oct 1;16(10):1453–61.
115. Penn DC, Povinelli DJ. On the lack of evidence that non-human animals possess anything remotely resembling a “theory of mind.” *Philos Trans R Soc B Biol Sci*. 2007 Apr 29;362(1480):731–44.
116. Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: The origins of cultural cognition. *Behav Brain Sci*. 2005 Oct;28(05):675–91.
117. Clayton NS, Dally JM, Emery NJ. Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philos Trans R Soc B Biol Sci*. 2007 Apr 29;362(1480):507–22.
118. Gweon H, Dodell-Feder D, Bedny M, Saxe R. Theory of Mind Performance in Children Correlates With Functional Specialization of a Brain Region for Thinking About Thoughts. *Child Dev*. 2012;83(6):1853–68.
119. Mitchell JP, Banaji MR, Macrae CN. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*. 2005 Dec;28(4):757–62.
120. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006 Apr;7(4):268–77.
121. 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*. 2001 Sep 28;293(5539):2425–30.
122. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*. 2006 Sep;10(9):424–30.
123. Koster-Hale J, Saxe R, Dungan J, Young LL. Decoding moral judgments from neural representations of intentions. *Proc Natl Acad Sci*. 2013 Mar 11;201207992.
124. Koster-Hale J, Bedny M, Saxe R. Thinking about seeing: perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition*. 2014 Oct;133(1):65–78.
125. Koster-Hale J, Richardson, Hilary, Saxe R. Mentalizing regions represent continuous, abstract dimensions of others’ beliefs. in prep;

126. Peelen MV, Atkinson AP, Vuilleumier P. Supramodal Representations of Perceived Emotions in the Human Brain. *J Neurosci*. 2010 Jul 28;30(30):10127–34.
127. Skerry, Amy, E., Saxe R. A common neural code for perceived and inferred emotions. *J Neurosci*. in press;
128. Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*. 1983 Jan;13(1):103–28.
129. Baillargeon R, Scott RM, He Z. False-belief understanding in infants. *Trends Cogn Sci*. 2010 Mar;14(3):110–8.
130. Saxe R. The New Puzzle of Theory of Mind Development. In: Banaji MR, Gelman SA, editors. *Navigating the Social World* [Internet]. Oxford University Press; 2013 [cited 2014 Oct 14]. p. 107–12. Available from: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199890712.001.0001/acprof-9780199890712-chapter-20>
131. Scholl BJ, Leslie AM. Modularity, development and “theory of mind.” *Mind Lang*. 1999;14(1):131–53.
132. Carlson SM, Moses LJ, Claxton LJ. Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *J Exp Child Psychol*. 2004 Apr;87(4):299–319.
133. Sabbagh MA, Xu F, Carlson SM, Moses LJ, Lee K. The Development of Executive Functioning and Theory of Mind. *Psychol Sci*. 2006 Jan;17(1):74–81.
134. Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci*. 2013 Oct 8;110(41):16616–21.
135. Sabbagh MA, Bowman LC, Evraire LE, Ito JMB. Neurodevelopmental Correlates of Theory of Mind in Preschool Children. *Child Dev*. 2009 Jul 1;80(4):1147–62.
136. Saxe RR, Whitfield-Gabrieli S, Scholz J, Pelphrey KA. Brain Regions for Perceiving and Reasoning About Other People in School-Aged Children. *Child Dev*. 2009;80(4):1197–209.
137. Apperly IA, Butterfill SA. Do humans have two systems to track beliefs and belief-like states? *Psychol Rev*. 2009 Oct;116(4):953–70.
138. Kovács ÁM, Téglás E, Endress AD. The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*. 2010 Dec 24;330(6012):1830–4.
139. Low J, Perner J. Implicit and explicit theory of mind: State of the art. *Br J Dev Psychol*. 2012 Mar 1;30(1):1–13.
140. Onishi KH, Baillargeon R. Do 15-Month-Old Infants Understand False Beliefs? *Science*. 2005 Apr 8;308(5719):255–8.

141. Southgate V, Senju A, Csibra G. Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci.* 2007 Jul;18(7):587–92.
142. Kovacs AM, Kuhn S, Gergely G, Csibra G, Brass M. Are All Beliefs Equal? Implicit Belief Attributions Recruiting Core Brain Regions of Theory of Mind. *PLoS ONE* [Internet]. 2014 Sep 26 [cited 2014 Oct 14];9(9). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4178016/>
143. Schneider D, Slaughter VP, Becker SI, Dux PE. Implicit false-belief processing in the human brain. *NeuroImage* [Internet]. [cited 2014 Jul 22]; Available from: <http://www.sciencedirect.com/science/article/pii/S1053811914005862>
144. Döhnell K, Schuwerk T, Meinhardt J, Sodian B, Hajak G, Sommer M. Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *NeuroImage.* 2012 Apr 15;60(3):1652–61.
145. Gallese V, Goldman A. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci.* 1998 Dec 1;2(12):493–501.
146. Stich S, Nichols S. Folk Psychology: Simulation or Tacit Theory. *Mind Lang.* 1992;7(1):35–71.
147. Gallese V, Sinigaglia C. What is so special about embodied simulation? *Trends Cogn Sci.* 2011 Nov;15(11):512–9.
148. Meltzoff AN. “Like me”: a foundation for social cognition. *Dev Sci.* 2007 Jan 1;10(1):126–34.
149. Bedny M, Pascual-Leone A, Saxe RR. Growing up blind does not change the neural bases of Theory of Mind. *Proc Natl Acad Sci.* 2009 Jul 7;106(27):11312–7.
150. Pyers J, Villiers PA de. Theory of mind in deaf children. In: Baron-Cohen S, Lombardo M, Tager-Flusberg H, editors. *Understanding Other Minds* [Internet]. Oxford University Press; 2013 [cited 2014 Oct 15]. p. 344–63. Available from: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199692972.001.0001/acprof-9780199692972-chapter-19>
151. Laurence S, Margolis E. Concepts and Cognitive Science. In: Margolis E, Laurence S, editors. *Concepts: Core Readings.* MIT; 1999. p. 3–81.
152. Murphy GL, Medin DL. The role of theories in conceptual coherence. *Psychol Rev.* 1985;92(3):289–316.
153. Baker CL, Saxe R, Tenenbaum JB. Action understanding as inverse planning. *Cognition.* 2009 Dec;113(3):329–49.
154. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to Grow a Mind: Statistics, Structure, and Abstraction. *Science.* 2011 Mar 11;331(6022):1279–85.