



Annual Review of Psychology

The Acquisition of Person Knowledge

Stefano Anzellotti and Liane L. Young

Department of Psychology, Boston College, Boston, Massachusetts 02467, USA;
email: stefano.anzellotti@bc.edu, liane.young@bc.edu

Annu. Rev. Psychol. 2020.71. Downloaded from www.annualreviews.org
Access provided by Boston College on 10/04/19. For personal use only.

Annu. Rev. Psychol. 2020. 71:22.1–22.22

The *Annual Review of Psychology* is online at
psych.annualreviews.org

<https://doi.org/10.1146/annurev-psych-010419-050844>

Copyright © 2020 by Annual Reviews.
All rights reserved

Keywords

person knowledge, social perception, social inference, theory of mind, identity

Abstract

How do we learn what we know about others? Answering this question requires understanding the perceptual mechanisms with which we recognize individuals and their actions, and the processes by which the resulting perceptual representations lead to inferences about people's mental states and traits. This review discusses recent behavioral, neural, and computational studies that have contributed to this broad research program, encompassing both social perception and social cognition.



Contents

INTRODUCTION 22.2

PERSON PERCEPTION 22.2

 Recognizing Agents and Conspecifics 22.3

 Recognizing Individuals 22.5

 Recognizing Facial Expressions 22.6

 Recognizing Actions 22.7

PERSON KNOWLEDGE 22.8

 Emotions 22.8

 Beliefs and Intentions 22.10

 Traits 22.11

OUTSTANDING QUESTIONS 22.13

 How Is Person Knowledge Acquired from Observation? 22.13

 Toward Integrated Person Models 22.13

 Person Model Impairments 22.13

CONCLUSIONS 22.14

INTRODUCTION

The past decade has seen significant progress in the study of person perception and the representation of person knowledge. New methods for generating stimuli, analyzing brain data, and modeling behavior have led to new observations and opened the door to new questions. However, the literature on perception and the literature on person knowledge have remained largely separate, with limited interchange between them.

Building on recent advances, we can begin to envision the goal of understanding how people construct and use models of other agents, starting from the perceptual mechanisms that transform sensory inputs into representations of individuals and their actions; continuing with how these representations are used to infer emotions, beliefs, and traits; and concluding with how these inferences are used to understand and predict others’ behavior (**Figure 1**). These processes are deeply interrelated: Each of them depends on the inputs it receives from the others and on the behavioral functions it needs to support. In this article, we provide an overview of current research in this area, bringing together the literatures on social perception and person knowledge.

PERSON PERCEPTION

Perception plays a fundamental role in the acquisition of person knowledge. Observing others’ actions unfold in the world, we can make inferences about their emotions, beliefs, and traits. Even when we learn something about a person from a third party, the original observer must have engaged with the challenge of starting from a sequence of observations to infer a mental state or a trait that could then be communicated.

The term person perception has sometimes been used in the literature to refer to a variety of processes, including some that have little to do with perception itself (e.g., retrieving knowledge about a person from memory given their name). In this article, we reserve the term for the recognition of agents and their identities, and for the recognition of expressions and actions.

Person perception:

the process of recognizing the appearance and the actions of a person

Person knowledge:

information associated with a person, such as their mental states and traits

Agent: an object that can initiate actions



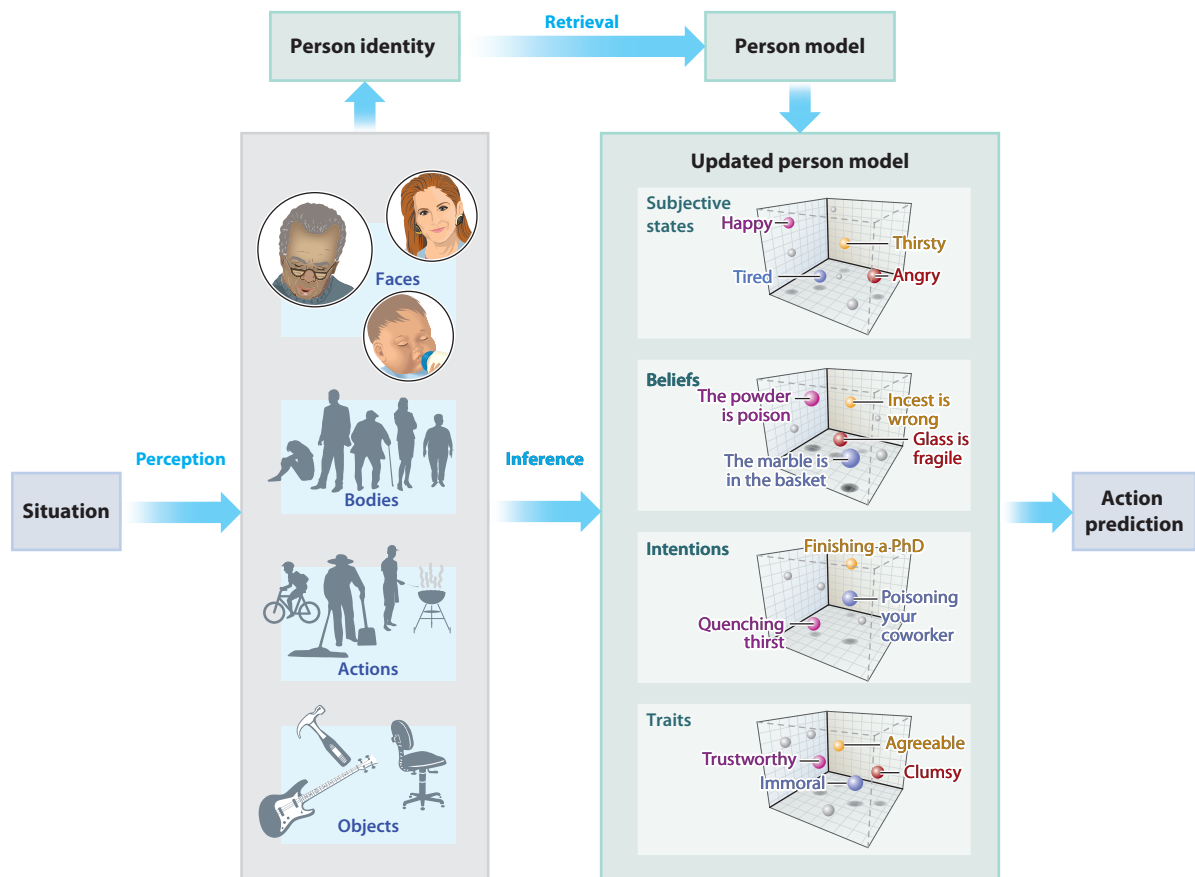


Figure 1

Schematic of the mechanisms engaged in the understanding and prediction of other people. Sensory inputs produced by a situation involving a person are mapped onto perceptual representations. Recognition of the identity of the person is used to retrieve the prior state of the person model for that identity. The perceptual representations are used to update the person model. The updated person model can be used to generate action predictions and can be modified if those predictions are violated. Note that the separate spaces for subjective states, beliefs, intentions, and traits are meant only to exemplify the variety of latent variables in a person model; we make no claim that these are different natural kinds with specialized neural mechanisms.

Recognizing Agents and Conspecifics

Recognizing an entity as an agent and as a conspecific comprises two fundamental steps in social cognition. They can lead an observer to attribute goals, traits, and beliefs and to expect that the entity might initiate actions. Furthermore, they can induce the observer to consider the entity's possible reactions to their own behavior.

Recognition of agents from static images. Entities can be visually recognized as agents through the use of static information such as shape or other cues (e.g., color, texture). Humans can detect the presence of animals and faces in static images rapidly and from very brief exposures. Above-chance animal detection is achieved for images presented for as short as 20 ms, and behavioral responses are produced as early as 290 ms after stimulus onset (Thorpe et al. 1996). The duration

FFA: fusiform face area

pSTS: posterior superior temporal sulcus

ASD: autism spectrum disorders

of 290 ms includes the time to plan and execute a motor response: Electroencephalography (EEG) data indicate that differences between the neural responses to animals and inanimate objects emerge as early as 150 ms post stimulus onset (Thorpe et al. 1996). It has been argued (Serre et al. 2007) that the speed of categorization suggests that this process is largely feed-forward.

The ability to discriminate between animate and inanimate objects from pictures is relatively robust to damage. Object recognition deficits affect the ability to categorize objects at a basic and subordinate level (Caramazza & Mahon 2003) but do not usually affect the ability to recognize whether or not an object is an animal. The animate–inanimate distinction is a large-scale principle of organization of the visual cortex (Chao et al. 1999). Category-specific brain regions showing selective responses to faces (Sergent et al. 1992, Kanwisher et al. 1997) and bodies (Downing et al. 2001) lie within broader areas of selectivity for animals, situated between the object-selective regions in the medial fusiform gyrus and the dorsal stream (Konkle & Caramazza 2013). Category specificity might be the outcome of computational demands (Leibo et al. 2015).

Brain regions showing selectivity for faces and bodies do not respond exclusively to conspecifics. For example, the fusiform face area (FFA) responds equally strongly to faces of humans and of cats (Tong et al. 2000). Despite this finding, several lines of evidence indicate that conspecifics hold a special status among types of animals. A recent study in humans found that human faces could be discriminated from animals in magnetoencephalography signals as early as 100 ms post stimulus onset (Cauchoix et al. 2014), suggesting that conspecifics might hold a special status among basic-level categories. Neuroimaging studies have found that faces of conspecifics can be discriminated from faces of other animal species on the basis of response patterns in the ventral prefrontal cortex (Anzellotti & Caramazza 2014a).

Motion, agency, and animacy. Motion cues play a critical role in the recognition of agents. Humans attribute goals and intentions even to geometric shapes that appear to move intentionally (Heider & Simmel 1944), suggesting that motion information is used not only to recognize known types of agents but also to infer that novel, never-before-encountered entities might be agents.

Biological motion selectively activates the posterior superior temporal sulcus (pSTS) (Pelphrey et al. 2005). Identifying regions distinguishing humans from other animals on the basis of biological motion is challenging: Humans are bipedal, while animals often are not, so pedalism can be a confound. A recent study used point-light displays depicting the motion of infants and chickens to control for pedalism and found that information distinguishing point-light displays of humans from animals across pedalism [i.e., for both infants and adults (Papeo et al. 2017)] could be decoded in the left pSTS and posterior cingulate. The right pSTS encodes information distinguishing bipeds from quadrupeds, but in the right pSTS responses to point-light displays of humans and of other bipedal animals (chickens) could not be distinguished (Papeo et al. 2017).

Geometric shapes that appear to move intentionally (Heider & Simmel 1944) elicit activity in the pSTS bilaterally and in lateral portions of the fusiform gyrus approximately corresponding to the areas that show increased responses to animals (Chao et al. 1999, Konkle & Caramazza 2013). This finding suggests that dynamic cues can induce objects of arbitrary shapes to be processed by neural systems specialized for animal recognition. More broadly, it suggests that the organization of ventral temporal regions might not be driven solely by static visual features. Importantly, biological motion recognition is impaired in children with autism spectrum disorders (ASD) (Blake et al. 2003); however, the causal link between deficits for biological motion and other social deficits in ASD remains unknown.

Agency is distinct from animacy: Self-initiated action is not unique to animals. Natural phenomena such as wind, rain, and avalanches have agency without animacy (Lowder & Gordon 2015). A recent study (Jozwik et al. 2018) used a wide array of stimuli to investigate the extent to

which animacy judgments are predicted by a set of relevant properties: “being alive,” “looking like an animal,” “having mobility,” “having agency,” and “being unpredictable.” The properties “being alive” and “having agency” were the most correlated, with correlation values $r > 0.6$. Attribution of animacy differs across cultures. For example, indigenous Ngäbe of Panama are more likely than US citizens to attribute intentions to plants and to infer that they engage in social actions such as kin altruism (Ojaletto et al. 2017).

ATL:
anterior temporal lobe

Recognizing Individuals

Social behavior relies critically on the recognition of people’s identity. Upon recognizing the identity of a person, we can acquire knowledge about them and then retrieve it on future encounters. Recognition of person identity relies on specialized neural mechanisms that can be selectively impaired while sparing the recognition of other object domains (Hecaen & Angelergues 1962, Rezlescu et al. 2014). Face-selective regions encode information about individual faces (Kriegeskorte et al. 2007, Natu et al. 2008, Nestor et al. 2011) and exhibit different extents of generalization across image transformations: The occipital face area (OFA) and FFA generalize only across changes in viewpoint (Anzellotti et al. 2013, Anzellotti & Caramazza 2014b), while the anterior temporal lobe (ATL) also generalizes across face parts (Anzellotti & Caramazza 2015).¹ A similar organization consisting of selective patches with different degrees of generalization has been observed in macaque monkeys (Freiwald & Tsao 2010). Furthermore, single neurons appear to represent dimensions of a face space and have orthogonal subspaces of faces within which their responses are approximately constant (Chang & Tsao 2017).

Person identity can also be recognized using information about someone’s face movements, gait, and voice. Observers use information about face movements during identity recognition (Dobs et al. 2016) and integrate it with shape information by weighting different cues depending on their reliability (Dobs et al. 2017). The identity of moving individuals can be decoded from the STS and, when the individuals are close to the observer, from body-selective regions such as the extrastriate body area and the fusiform body area (Hahn et al. 2016, Hahn & O’Toole 2017). Voice identity can also be decoded from the STS (Formisano et al. 2008, Anzellotti & Caramazza 2017, Hasan et al. 2016). Furthermore, the STS encodes representations of person identity that generalize across the visual and auditory modalities (Anzellotti & Caramazza 2017).

In addition to perceiving cues such as the face, body, movements, and voice, we perceive the context in which we encounter people—including the place and the moment in time in which we encounter them (Bar 2004, Bar et al. 2006, Kaiser & Cichy 2018, Castelhana & Pereira 2018). Faces presented in a context in which they had previously been seen are recognized faster (Hanczakowski et al. 2015). Furthermore, presenting novel faces in a context where another face had previously been shown increases false-alarm rates for judgments of face familiarity (Gruppuso et al. 2007).

Familiar faces and people. Familiar faces are recognized more accurately than unfamiliar faces using part of the face, and they are recognized more accurately through presentation of the inner part of the face versus the contour (Ellis et al. 1979). Familiar faces are also recognized faster than unfamiliar faces (Ramon et al. 2011). Familiar people are recognized more accurately than unfamiliar people in noisy videos showing their faces and bodies (Burton et al. 1999), and performance significantly worsens for familiar people when their faces are obscured. This finding suggests that information about the face contributes importantly to recognition and that person recognition

¹Different temporal stages of face processing have recently been identified with EEG (Kietzmann et al. 2017).



MPFC: medial prefrontal cortex

cannot be completed at the same level of accuracy when relying only on the body. Taken together, these observations show that familiarity alters the process of face recognition, leading to greater speed, accuracy, and robustness to information loss in the stimuli.

Familiar faces lead to stronger responses than do unfamiliar faces in several cortical regions, including the posterior cingulate, medial prefrontal cortex (MPFC), anterior STS, and hippocampus (Leveroni et al. 2000, Gobbini & Haxby 2007). Recognition of familiar faces and matching of unfamiliar faces show double dissociations in patients (Malone et al. 1982, Young et al. 1993). A patient with damage to the medial parietal regions demonstrated impaired recognition of familiar and famous people, alongside spared matching of different images of unfamiliar faces with the same identity (Malone et al. 1982). This finding is in line with the possible causal role of the medial parietal cortex (i.e., posterior cingulate) in familiar face recognition.

Patients with semantic dementia (Hodges et al. 1992) can also present with deficits for the recognition of famous people, with greater deficits for the recognition of famous faces in patients with disproportionate atrophy to the right hemisphere and greater deficits for the recognition of famous names in patients with disproportionate atrophy to the left hemisphere (Snowden et al. 2004). These observations and more recent neuroimaging results (Wang et al. 2017) have led to the proposal that the ATL might serve as a hub for the integration of knowledge about people (Wang et al. 2017). In macaques, familiar faces disproportionately activate a patch in the temporal pole and another in the entorhinal cortex (Landi & Freiwald 2017). The relative contributions of the posterior cingulate and ATL remain unknown.

Recognizing Facial Expressions

Facial expressions can provide important cues about others' mental states. The view that expressions are clear and unambiguous indicators of specific emotions (Ekman 1992, 1999), however, is challenged by several lines of evidence showing that facial expressions can be highly ambiguous (Aviezer et al. 2012), demonstrating the importance of context for expression recognition (Carroll & Russell 1996, Barrett et al. 2011, Hassin et al. 2013).

However, spontaneous facial expressions still convey sufficient information to infer emotions with above-chance accuracy² (Wagner et al. 1986), and expressions can be used to infer desires and beliefs (Wu & Schulz 2018). Furthermore, facial expressions might be used to disambiguate between alternative possible emotional reactions to a given context (Saxe & Houlihan 2017). Facial expression recognition is affected by familiarity: Expressions of famous people are recognized more accurately than expressions of unfamiliar people (Baudouin et al. 2000).

An influential account of the neural bases of face processing holds that facial expressions and face identity are processed by distinct pathways (Haxby et al. 2000). In this account, the ventral temporal regions (OFA and FFA) are specialized for face identity, while the pSTS is specialized for facial expressions. Indeed, in the pSTS, emotional faces yield a stronger response than neutral faces (Engell & Haxby 2007) and produce response patterns that can be used to decode emotional valence (Skerry & Saxe 2014). Multivariate analyses show that the pSTS encodes information about emotion that generalizes across facial expressions, voices, and body posture (Peelen et al. 2010, Skerry & Saxe 2014), as well as information about specific face movements (Srinivasan et al. 2016, Deen & Saxe 2019).

However, support for a complete separation between identity recognition and expression recognition is weaker than is often assumed (Calder & Young 2005, Bernstein & Yovel 2015).

²Here a participant's response is considered accurate if it matches the consensus of independent observers. This is to some extent an abuse of language because the true emotion that was experienced by the person in the image is usually not known.

Emotional faces yield stronger responses than neutral faces not only in the pSTS but also in the occipital and fusiform regions (Engell & Haxby 2007), and the valence of expressions can be decoded from the OFA and the FFA (Skerry & Saxe 2014). In addition, recent results show that face identity can be decoded from patterns of activity in the pSTS (Anzellotti & Caramazza 2017), and a patient with a lesion involving the pSTS (Fox et al. 2011, patient 5) showed impairments not only in the recognition of facial expressions but also in the recognition of identity across different expressions.

A possible explanation for the finding that identity can be decoded from the pSTS is that information about identity is not discarded entirely in the pathway for recognition of facial expressions. According to an alternative hypothesis, it might be computationally efficient to implement the recognition of identity and expressions within the same neural mechanisms: Recognition of identity could help isolate which aspects of an image are due to expression, and vice versa.³

If recognition of identity and of expression are intertwined, we would predict that as recognition of facial expressions improves from region to region in the processing hierarchy, recognition of identity would also improve. A recent study found that classification of identity using features from the layers of a deep network trained to label facial expressions increased from layer to layer, even though the deep network had not been trained to recognize identity (K.C. O'Neil, R.R. Saxe and S. Azellotti, manuscript in preparation).

The broader area surrounding the face-selective pSTS might play a more general role in recognition of identity from face and body motion. Gait can be used not only for recognizing conspecifics (Papeo et al. 2017) but also for recognizing identity (O'Toole et al. 2011, Simhi & Yovel 2016). Regions of the pSTS neighboring the face-selective pSTS respond to biological motion and to point-light displays (Martin & Weisberg 2003). Taken together, these findings suggest that a larger patch of the pSTS, including the face-selective pSTS, might be a multimodal convergence zone integrating motion and form information as well as auditory information (Yovel & O'Toole 2016, Peelen et al. 2010, Anzellotti & Caramazza 2017).

Recognizing Actions

Recognizing the actions of other agents is critical for detecting threats, engaging in cooperation, and coordinating our own actions with consideration for the social context around us. Actions can be recognized at different levels of abstraction. For example, transitive actions (actions that involve an object) can be categorized according to the category of the object involved: We can recognize the opening or closing of a specific bottle, of any bottle, or of any object [e.g., a bottle and a box (Wurm & Lingnau 2015)]. Actions can also be categorized at different levels of abstraction according to their goal. We can recognize the action of clapping or, at a more abstract level, the action of producing sound.

Neural mechanisms for action recognition. Neuroimaging studies show that action observation leads to increased responses in the lateral occipito-temporal cortex (LOTc) (Watson et al. 2013, Lingnau & Downing 2015), as well as the pSTS, anterior intraparietal sulcus/inferior parietal lobule (aIPS/IPL), ventral premotor cortex (PMv), and supplementary motor area (Grafton et al. 1996, Rizzolatti et al. 1996, Buccino et al. 2001, Molnar-Szakacs et al. 2006, Cross et al. 2006). This network of brain regions is often referred to as the action observation network (Calvo-Merino et al. 2006, Cross et al. 2009). A similar network of brain regions is activated by biological motion (Grèzes et al. 2001).

³This hypothesis is consistent with the proposal that the pSTS might also contribute to recognition of identity from dynamic stimuli (O'Toole et al. 2002).



Furthermore, different actions can be decoded from the patterns of response in the LOTC (Oosterhof et al. 2010), aIPS/IPL (Dinstein et al. 2008), and PMv (Wurm & Lingnau 2015). In the PMv, decoding of actions succeeds at the most concrete level, but generalization (e.g., opening versus closing across different types of bottles, or opening versus closing of bottles and boxes) fails (Wurm & Lingnau 2015). Furthermore, decoding in the PMv succeeds only when participants are explicitly requested to recognize actions (Wurm et al. 2016). By contrast, in the aIPS/IPL and LOTC, classification succeeds at both concrete and abstract levels (Wurm & Lingnau 2015), even when recognizing actions is not required by the task (Wurm et al. 2016). Representations of actions in portions of the LOTC generalize across videos and sentences (Wurm & Caramazza 2018), lending additional support to the view that the LOTC encodes abstract representations of actions. Representations in the aIPS/IPL encode information about abstract functions of objects (e.g., “an umbrella is for protecting oneself from the rain”). Whether such representations and the aIPS/IPL representations of actions (Wurm & Lingnau 2015, Wurm et al. 2016) overlap remains unknown.

Recent research has shown that the LOTC is organized at a macroscopic scale by the transitivity and sociality of actions. Pattern similarity in dorsal portions of the LOTC reflects how similar actions are in terms of sociality, whereas pattern similarity in ventral portions reflects how similar actions are in terms of transitivity (Wurm et al. 2017).

According to an influential proposal, recognizing actions relies on the neural mechanisms for action execution (Rizzolatti et al. 2001, Rizzolatti & Craighero 2004, Rizzolatti & Sinigaglia 2016): Action understanding consists of direct mapping from perception to motor representations of an action (Iacoboni et al. 1999). This perspective, however, is fraught with theoretical and empirical issues (Caramazza et al. 2014). The finding that mirror neurons respond to both observed actions and executed actions is symmetrical: It could just as well be used to claim that action execution is performed by visual simulation. Furthermore, observed actions are different from any action the observer can perform (they are performed with a different body); therefore, some abstraction would need to occur before the appropriate motor representations could be accessed. At the empirical level, there is extensive evidence for precisely this type of abstraction in the LOTC (Wurm & Lingnau 2015, Wurm et al. 2017), which does not respond during action execution. In addition, patients with impairments in action execution (Negri et al. 2007) and patients with upper-limb dysplasia (Vannuscorps & Caramazza 2016) can have spared action recognition.

PERSON KNOWLEDGE

Humans represent a wealth of information about others, ranging from current mental states (e.g., emotions and thoughts) to more lasting traits (e.g., personality and moral values), and from semantic knowledge (e.g., occupation) to episodic memories of particular moment spent with someone (e.g., meeting family arriving at the airport). Understanding person knowledge requires understanding (*a*) what information we represent about others, (*b*) how this information is acquired, and (*c*) how this information is used to make new inferences and decisions. In separate subsections we discuss emotions, beliefs and intentions, and traits, but this does not amount to a claim that they are distinct natural kinds. Because of space limitations, for semantic knowledge we refer the reader to recent reviews (Ralph et al. 2017, Leshinskaya et al. 2017, Yee et al. 2018).

Emotions

Since the focus of this article is on person knowledge, we discuss emotion attribution, not the first-person experience of emotions. For this reason, some important theories of first-person emotion recognition (e.g., Barrett 2014) are not discussed in detail.

Structure. The theory of basic emotions (Ekman 1992) proposes that emotions can be represented as a vectors of five values encoding the intensity of each of the basic emotions: anger, disgust, fear, happiness, and sadness. By contrast, according to the circumplex model, emotions are represented as lying on a circle within the space spanned by valence and arousal (Russell 1980, Barrett & Russell 1998, Russell & Barrett 1999, Russell et al. 2003). The distance between two emotions in this space reflects the similarity between them. A key idea introduced by the circumplex theory (Russell 1980) is that emotions might be best represented not as a vector space but rather as a manifold (e.g., Tenenbaum 1998).

Recent models suggest that more than five dimensions might be needed to capture human emotion attribution. A space consisting of 38 appraisal dimensions was found to outperform a model using five basic emotions and a model using valence and arousal (Skerry & Saxe 2015). An optimized 10-dimensional space could achieve a performance very similar to that of the 38 appraisal dimensions (Skerry & Saxe 2015). It remains unknown whether representations of emotions may be captured by an even lower dimensional nonlinear manifold embedded in this 10-dimensional space. The dimensionality of emotion space may also depend on the stimuli used: A broader range of stimuli might elicit a variety of emotions that require more dimensions.

Observers are usually uncertain about the emotion experienced by an agent. Emotion attribution can be thought as the process of inferring a probability distribution on the space of emotions (Gygax et al. 2003, Ong et al. 2015, Saxe & Houlihan 2017). The language of probability also helps differentiate among the notions of similarity, independence, and transition probability. We can use the term independence in the sense of probability theory: Two emotions are independent if knowing that a person is experiencing one of them does not affect how likely it is that the same person is experiencing the other. Two emotions might be dissimilar but not independent, and vice versa. For example, being surprised and being upset feel quite different, but someone who does not like surprises might often be upset when she is surprised (dependence without similarity). Finally, in addition to asking whether the presence of an emotion at a given time makes another emotion more or less likely at the same time, we can ask whether one emotion makes another emotion more or less likely some time later (transition probability).

In summary, representations of emotions could be thought of as consisting of an emotion space or manifold, equipped with (*a*) a similarity metric and (*b*) a stochastic process that captures the nonindependence between emotion dimensions and the dependence of emotions on their history (Lewis 2005, Thornton & Tamir 2017, Tamir & Thornton 2018). Individual-specific stochastic processes can be learned, representing information about the temporal dynamics of an emotion for that individual (e.g., “Does he hold a grudge?”) and the interactions between different emotions (e.g., “Does she get upset when she is surprised?”). These issues bring us closer to the topic of traits, which is discussed below.

Inference. Observers can infer emotions from several types of cues, including facial expressions, actions, and situations (Gygax et al. 2003, Skerry & Saxe 2014). Recent studies propose to understand emotion attribution using ideas from probability theory and Bayesian model inversion (Ong et al. 2015, Saxe & Houlihan 2017). These studies are part of a broader literature using Bayesian models as a window into several facets of person knowledge, including the attribution of desires (Baker et al. 2009, Baker & Tenenbaum 2014, Baker et al. 2017), intentions (Jern & Kemp 2015; Jara-Ettinger et al. 2016, 2017), and preferences (Jern et al. 2011, Gershman et al. 2017). In the case of emotions, partially observable causes (e.g., a situation) lead to nonobservable emotional states, which in turn lead to observable actions and facial expressions. As a consequence, the probability of nonobservable emotional states can be inferred on the basis of the observable causes and the observable actions and expressions, combining a model that links the observable causes to the

Bayesian model inversion: application of Bayes's theorem to make inferences about unobservable variables by using a model of how they lead to observable consequences



likely resulting emotional states and the inversion of a model that links the emotional states to the observed actions and facial expressions (Saxe & Houlihan 2017).

Neural bases. The neural bases for the recognition of facial expressions have been discussed in detail in the section titled Person Perception. Recent research investigated representations of emotions, when they are recognized from a facial expression, and when they are inferred on the basis of information about a situation (without any facial expressions shown) (Skerry & Saxe 2014). The dorsomedial prefrontal cortex (DMPFC) was found to encode the valence of emotions generalizing across facial expressions and situations (Skerry & Saxe 2014), suggesting that this brain region encodes abstract representations of the valence of emotions.

Beliefs and Intentions

A key component of person knowledge is the representation of others' beliefs and intentions. Understanding the beliefs and intentions of other agents is essential to accurately predict their future actions (Baron-Cohen et al. 1985).

Structure. Candidate theories of the representations of one's own beliefs can also be used as candidate theories of how we represent the beliefs of others.⁴ Modeling the structure of beliefs is extremely challenging. One challenge is that beliefs are tied to the complexity of the world, and the world itself is changing. Models in the field of artificial intelligence and natural language processing attempt to capture world knowledge as a network of concepts and their relations (Speer et al. 2017, Miller 1995, Goodman et al. 2014). Another challenge comes from the fact that different representational structures might be used as a function of the task: this view has been put forward in the context of "commonsense knowledge" (Minsky 2000). Improving models of world knowledge is a key direction of research in current artificial intelligence (Shi & Weninger 2018), and research in social psychology could leverage these advances to investigate how we represent the beliefs of others. In addition to these complexities, beliefs about other agents can be recursive ["I believe that she believes that I believe..."] (Goodie et al. 2012)], and models with many levels of recursion rapidly become intractable.

Several models have been proposed for the structure of intentions/goals (Austin & Vancouver 1996). The most influential accounts are based on a hierarchy of goals organized in clusters at different levels (e.g., Chulef et al. 2001). Dimensional accounts of goals have also been proposed (e.g., Winell 1987), leading to six proposed factors: "importance," "difficulty," "specificity," "temporal range," "level of consciousness," and "connectedness" (Austin & Vancouver 1996).

Inference. A wealth of research has investigated the development of intention understanding; for lack of space, we refer the reader to existing reviews (e.g., Tomasello et al. 2005). Recent studies have focused on the attribution of intentions and beliefs in controlled settings, where the space of beliefs and intentions is restricted so that it becomes tractable (Baker et al. 2009, 2017). Bayesian models have been successful at modeling human inferences about beliefs and intentions in these controlled settings, mirroring closely the inferences made by participants (Baker & Tenenbaum 2014, Baker et al. 2017). The framework of partially observable Markov decision processes (Cassandra 1998) has been used to account for how multiple sequential observations are integrated during inference (Baker & Tenenbaum 2014, Baker et al. 2017).

⁴Although, of course, the appropriate theories for the representation of one's own beliefs and for the representation of others' beliefs might be different.

Neural bases. A wealth of research has investigated the neural mechanisms by which humans attribute beliefs to others, consistently identifying a network of brain regions including the dorsal and ventral subregions of the MFPC (DMPFC, VMPFC), right and left temporoparietal junction (RTPJ, LTPJ), and precuneus. Brain regions in the theory of mind network (ToM) (Frith & Frith 2000, Gallagher & Frith 2003) show stronger responses when participants read stories about others' thoughts and feelings than when they read about physical properties of objects (Fletcher et al. 1995). The same effect holds when the stories are presented with visual vignettes in the absence of text (Gallagher et al. 2000). These regions also respond more when participants attribute false beliefs to a character than during control tasks such as inferring a physical process (e.g., melting or rusting) or representing false photographs or maps (Saxe & Kanwisher 2003); they also respond more to sentences describing thoughts versus facts (Zaitchik et al. 2010). A recent activation likelihood estimation meta-analysis of 144 datasets (3,150 participants) uncovered MPFC and bilateral TPJ activation across all ToM tasks sampled (Molenberghs et al. 2016). In summary, regions in the ToM respond during the attribution of beliefs to others across a variety of experimental paradigms (for an in-depth review, see Koster-Hale & Saxe 2013).

Patterns of activity in the RTPJ can be used to decode whether participants think that another's action was intentional or accidental (Koster-Hale et al. 2013) and to decode the strength of the evidence supporting a belief as well as the modality through which the belief was acquired [Koster-Hale et al. 2017; see also Mengotti et al. 2017 for transcranial magnetic stimulation (TMS) evidence that the RTPJ contributes to updating probabilistic beliefs]. Other recent research has combined inhibitory continuous theta-burst TMS with model-based functional magnetic resonance imaging (fMRI) to look at the causal role of the RTPJ in ToM in a game context (Hill et al. 2017). TMS to the RTPJ disrupted participants' estimation of how their own actions would influence the other players' strategy, as well as the functional connectivity of the RTPJ to the VMPFC and DMPFC.

Traits

In addition to mental states, beliefs, and intentions, observers attribute to others more lasting properties—traits (Allport & Odbert 1936). Whether or not states and traits are qualitatively different is an open issue in the literature (Allen & Potkay 1981; S. Anzellotti, manuscript in preparation).

Structure. Early research has led to the identification of five factors that capture most of participants' variability in trait ratings, where the ratings were collected with scales asking questions such as, "To what degree is person X fearful?" (Digman 1990). The finding of five reliable factors has been replicated by several groups of researchers (Tupes & Christal 1992, Norman 1963) and across very different populations of participants, such as teachers rating children and college students rating one another (Digman & Takemoto-Chock 1981).

Other dimensions capturing traits that have been proposed in the literature include warmth and competence (Fiske et al. 2002), agency and experience (Gray et al. 2007), and trustworthiness and social dominance (Oosterhof & Todorov 2008). A recent study collected behavioral ratings along the dimensions proposed by previous theories for 60 famous people chosen to span a variety of traits, and used principal component analysis to identify three dimensions that explain most of the variance ("power," "valence," and "sociality") (Thornton & Mitchell 2017). These dimensions account for a high percentage (66%) of the reliable variance in fMRI responses within regions responding reliably to the 60 people's names (Thornton & Mitchell 2017).

RTPJ: right temporoparietal junction

LTPJ: left temporoparietal junction

ToM: theory of mind network

TMS: transcranial magnetic stimulation

Updating: integrating new observations with a stored person model to compute a revised person model

fMRI: functional magnetic resonance imaging

More recently, it has been proposed that other people might be represented as the sum of the mental states they usually experience (Thornton et al. 2018), and this so-called sums of states model has been shown to outperform the previous three-dimensional model (Thornton & Mitchell 2017) at explaining neural responses (Thornton et al. 2018). However, the sums of states model accounts only for the frequency of mental states and not for individual differences in responding to different situations. For instance, two different individuals might experience fear equally often, but one could be afraid of heights while the other could be afraid of spiders—it seems unlikely that human observers would not represent the differences between these individuals.

Inference. Behavioral studies have investigated how participants form representations of others based on descriptions, behaviors (Hastie 1980), nonverbal behaviors (Kraft-Todd et al. 2017), and face images (Todorov & Uleman 2002, 2003). A seminal paper (Hastie 1980) introduced a model of the mechanism of impression formation and of the retrieval of person knowledge that was subsequently expanded to account for a variety of behavioral findings on person memory (Srull & Wyer 1989).

Recently, studies on intelligence attribution (Kryven et al. 2016, Kryven 2018) have used controlled experimental paradigms, Bayesian models, and the inverse planning approach (Baker et al. 2009) to study how observers judge the ability of players weighing the importance of successful outcomes versus optimal strategies. This research takes initial steps toward the long-term goal of building task-performing models that can produce human-like trait inferences based on perceptual inputs.

Neural bases. The section titled Familiar Faces and People, above, describes the involvement of the posterior cingulate, anterior STS and hippocampus, and MPFC. Early social neuroscience studies found stronger responses in the MPFC when participants were presented with behaviors and asked to form an impression about the character performing those behaviors than when they were presented with the same behaviors and asked to remember their order (Mitchell et al. 2004).

Furthermore, the MPFC responds more during trials that will be successfully remembered (Mitchell et al. 2004, Baron et al. 2010) and during formation of impressions about people than during formation of impressions about objects (Mitchell et al. 2005). In addition, the MPFC responds more to actions that are diagnostic of traits than to actions that are not (e.g., “he played his music loud at the public picnic grounds” versus “he ordered a cup of coffee at Starbucks”) (Mitchell et al. 2006). Response patterns in the MPFC also distinguish between individuals with high and low agreeableness, and between individuals with different trait combinations (Hassabis et al. 2013).

After a participant is told a set of behaviors implying a trait, a new inconsistent behavior leads to stronger responses in the MPFC (Ma et al. 2011, Mende-Siedlecki et al. 2012). A recent study (Ferrari et al. 2016) found that applying TMS to the MPFC affects impression updating, reducing the extent to which participants revise their judgments of an individual from trustworthy to untrustworthy. This finding provides causal evidence for the involvement of the MPFC in trait representations. We refer the reader to a recent review (Mende-Siedlecki 2018) for an in-depth discussion of impression updating.

In addition to the MPFC, other brain regions might be involved in the representation of traits. Individuals high versus low in extraversion can be classified from the posterior cingulate, but not from the MPFC (Hassabis et al. 2013). Furthermore, a recent study (Spiers et al. 2017) found that the ATL tracks the valence associated with social groups. A wealth of neuropsychological studies implicates the ATL in the representation of semantic knowledge about people (Ellis et al. 1989, Snowden et al. 2012); its role in the representation of traits remains to be elucidated.

OUTSTANDING QUESTIONS

The past decade has seen significant progress in our understanding of the acquisition of person knowledge. Despite this progress, many questions remain open.

How Is Person Knowledge Acquired from Observation?

The study of how knowledge is acquired from perceptual inputs is in its infancy, and it is challenging in several respects. First, it requires a joint understanding of the perceptual mechanisms providing the inputs from inference and of the inferential processes (Schirmer & Adolphs 2017, Grill-Spector et al. 2018). Second, it requires consideration of the integration of observations about an agent and about context (Carroll & Russell 1996, Aviezer et al. 2012, Saxe & Houlihan 2017, Baker et al. 2017). Third, it requires us to characterize the acquisition of information over time, as well as the use of previously acquired knowledge jointly with new perceptual inputs (Ma et al. 2011, Mende-Siedlecki et al. 2012, Hassabis et al. 2013). Current computational studies have focused mostly on states, and computational models of trait learning are few [intelligence is a notable exception (Kryven et al. 2016, Kryven 2018)]. Furthermore, existing models are usually limited to constrained scenarios, and cannot explain the acquisition of person knowledge in naturalistic settings.

Toward Integrated Person Models

Most studies in the literature adopt a divide and conquer strategy, isolating a particular kind of representation or inference. However, understanding other agents requires person models that capture the interactions and causal relations among emotions, goals, beliefs, traits, and other properties of an agent.

The need for rich models that capture multiple aspects of an agent and go beyond simple statistical associations is not new. The literature on schemas in the 1980s arose from the realization that associative models were inadequate for a variety of cognitive processes (Simon 1978, Anderson 1980, Hastie 1980). At the time, formalizing these latent models and using them to generate measurable predictions proved a daunting challenge (Fiske & Linville 1980).

Current computational techniques are beginning to make this kind of research possible. The investigation of individual types of representations and processes is leading to the study of pairwise interactions in the recognition of identity and expressions (Dobs et al. 2016; K.C. O'Neill, R.R. Saxe and S. Anzellotti, manuscript in preparation), of goals and beliefs (Baker & Tenenbaum 2014), of states and traits (Tamir & Thornton 2018; S. Anzellotti, manuscript in preparation), and of group membership and moral judgments (Waytz & Young 2018).

A related issue is that observers likely use different models at different levels of complexity in different circumstances (Minsky 2000, Gershman et al. 2016). Therefore, understanding how humans acquire knowledge about others likely means understanding not only one model of agents but possibly an ensemble of models, together with mechanisms to select which models in the ensemble to adopt depending on the situation.

Person Model Impairments

Finally, disorders of social cognition such as ASD affect millions of people across the world (Christensen et al. 2018). Our understanding of the cognitive and neural mechanisms affected in these disorders is still very limited. We hope that the formulation of models that capture the inferences and predictions of healthy controls and the study of their neural implementation can

Person model: a model of an individual consisting of a set of variables (i.e., states and traits, as well as physical properties such as strength and height) and causal relations between them that can be used to generate action predictions and that can be updated by comparing the predictions with the observations



help determine quantitatively which components of the computations are affected in patients, and lead to a clearer picture of the underlying neural impairments.

CONCLUSIONS

In this article, we have attempted to collect an overview of studies ranging from perception to social cognition that are converging to shape our understanding of how humans recognize and make sense of others. We have argued that this literature can be unified under the broad research program of understanding how humans learn, represent, and use models of other agents. Different parts of this research program are being pursued by distinct communities of scientists, and interactions between them are often limited.

A key goal for the future of the field is the construction of models of other agents that can match the human ability to predict the behavior of other agents (see Kriegeskorte & Douglas 2018 for a broader discussion of task-performing models). These models will need to infer unobservable latent variables such as emotions, beliefs, goals, and traits from observations of agents' behavior, and will need to use these latent variables to generate human-like predictions of the agents' future actions. Furthermore, these models will need to support human-like decisions and judgments (e.g., in the moral domain). Moving toward this goal will require greater interaction and communication between different communities of researchers.

As in most domains of psychology and neuroscience, direct measurement of the mechanisms implementing the computations underlying the acquisition of person knowledge at the resolution and scale needed to reconstruct them artificially is currently beyond our reach. However, we can observe their traces in behavior and in different brain measures. A critical future direction will involve jointly leveraging behavioral and neural data across multiple methods, as if they were shades on a wall from which we need to recover the object that is casting them.

SUMMARY POINTS

1. Humans use observation to acquire knowledge about other people's mental states and traits.
2. Investigating how humans acquire and use this knowledge requires convergence between the literature on perception and the literature on social cognition.
3. Perceptual mechanisms enable recognition of agents and their actions, providing the necessary inputs for inference.
4. Perceptual inputs are used to infer identity-specific person models with which observers can predict others' actions.
5. Recognition of agents and their actions is implemented by networks of specialized brain regions encoding representations at multiple levels of abstraction (OFA, FFA, STS, and ATL for faces; the extrastriate body area and fusiform body area for bodies; LOTC, aIPS, and PMv for actions).
6. ToM regions, including the TPJ, MPFC, and precuneus, support person knowledge.
7. Developing task-performing computational models of the acquisition of person knowledge jointly constrained by behavioral and neural data is a promising direction for future research.

FUTURE ISSUES

1. What are the computations supporting the recognition of agents?
2. How are perceptual representations of agents, objects, and actions integrated to make inferences about person knowledge?
3. What is the relationship between mental states (beliefs, intentions) and traits within a person model?
4. How can we jointly leverage behavioral and neural data in an optimal way to test task-performing models of person knowledge acquisition?
5. How is person knowledge acquired through observation integrated with knowledge communicated by language?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Josh Hirschfeld-Kroen for comments and assistance in the preparation of the manuscript, and Fiery Cushman, Max Kleiman-Weiner, James Russell, Peter Mende-Siedlecki, Lily Tsoi, BoKyung Park, Minjae Kim, Kathryn C. O'Neill, Mengting Fang, Kirstan Brodie, Ryan McManus, Justin Martin, Dominic Fareri, James Dungan, and Gordon Kraft-Todd for comments on earlier versions of the manuscript. The authors acknowledge the support of Boston College, the Simons Foundation, and the John Templeton Foundation.

LITERATURE CITED

- Allen BP, Potkay CR. 1981. On the arbitrary distinction between states and traits. *J. Personal. Soc. Psychol.* 41:916–28
- Allport GW, Odbert HS. 1936. Trait-names: a psycho-lexical study. *Psychol. Monogr.* 47:i–171
- Anderson JR. 1980. Concepts, propositions, and schemata: What are the cognitive units? *Tech. Rep., Dep. Psychol.*, Carnegie Mellon Univ., Pittsburgh, PA
- Anzellotti S, Caramazza A. 2014a. Individuating the neural bases for the recognition of conspecifics with mvpa. *NeuroImage* 89:165–70
- Anzellotti S, Caramazza A. 2014b. The neural mechanisms for the recognition of face identity in humans. *Front. Psychol.* 5:672
- Anzellotti S, Caramazza A. 2015. From parts to identity: invariance and sensitivity of face representations to different face halves. *Cereb. Cortex* 26:1900–9
- Anzellotti S, Caramazza A. 2017. Multimodal representations of person identity individuated with fMRI. *Cortex* 89:85–97
- Anzellotti S, Fairhall SL, Caramazza A. 2013. Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex* 24:1988–95
- Austin JT, Vancouver JB. 1996. Goal constructs in psychology: structure, process, and content. *Psychol. Bull.* 120:338–75
- Aviezer H, Trope Y, Todorov A. 2012. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338:1225–29



- Baker CL, Tenenbaum JB. 2014. Modeling human plan recognition using Bayesian theory of mind. In *Plan, Activity, and Intent Recognition: Theory and Practice*, ed. G Sukthankar, C Geib, HH Bui, DV Pynadath, RP Goldman, pp. 177–204. Amsterdam: Elsevier
- Baker CL, Saxe R, Tenenbaum JB. 2009. Action understanding as inverse planning. *Cognition* 113:329–49
- Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* 1:0064
- Bar M. 2004. Visual objects in context. *Nat. Rev. Neurosci.* 5:617–29
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, et al. 2006. Top-down facilitation of visual recognition. *PNAS* 103:449–54
- Baron SG, Gobbini MI, Engell AD, Todorov A. 2010. Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Soc. Cogn. Affect. Neurosci.* 6:572–81
- Baron-Cohen S, Leslie AM, Frith U. 1985. Does the autistic child have a “theory of mind”? *Cognition* 21(1):37–46
- Barrett LF. 2014. The conceptual act theory: a précis. *Emot. Rev.* 6:292–97
- Barrett LF, Mesquita B, Gendron M. 2011. Context in emotion perception. *Curr. Dir. Psychol. Sci.* 20:286–90
- Barrett LF, Russell JA. 1998. Independence and bipolarity in the structure of current affect. *J. Personal. Soc. Psychol.* 74:967–84
- Baudouin JY, Sansone S, Tiberghien G. 2000. Recognizing expression from familiar and unfamiliar faces. *Pragmat. Cogn.* 8:123–46
- Bernstein M, Yovel G. 2015. Two neural pathways of face processing: a critical evaluation of current models. *Neurosci. Biobehav. Rev.* 55:536–46
- Blake R, Turner LM, Smoski MJ, Pozdol SL, Stone WL. 2003. Visual recognition of biological motion is impaired in children with autism. *Psychol. Sci.* 14:151–57
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, et al. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13:400–4
- Burton AM, Wilson S, Cowan M, Bruce V. 1999. Face recognition in poor-quality video: evidence from security surveillance. *Psychol. Sci.* 10:243–48
- Calder AJ, Young AW. 2005. Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* 6:641–51
- Calvo-Merino B, Grèzes J, Glaser DE, Passingham RE, Haggard P. 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16:1905–10
- Caramazza A, Anzellotti S, Strnad L, Lingnau A. 2014. Embodied cognition and mirror neurons: a critical assessment. *Annu. Rev. Neurosci.* 37:1–15
- Caramazza A, Mahon BZ. 2003. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* 7:354–61
- Carroll JM, Russell JA. 1996. Do facial expressions signal specific emotions? Judging emotion from the face in context. *J. Personal. Soc. Psychol.* 70:205–18
- Cassandra AR. 1998. *A survey of POMDP applications*. Work. Pap., Microelectron. Comput. Technol. Corp., Austin, TX. <http://www.pomdp.org/papers/applications.pdf>
- Castelhano MS, Pereira EJ. 2018. The influence of scene context on parafoveal processing of objects. *Q. J. Exp. Psychol.* 71:229–40
- Cauchoix M, Barragan-Jason G, Serre T, Barbeau EJ. 2014. The neural dynamics of face detection in the wild revealed by MVPA. *J. Neurosci.* 34:846–54
- Chang L, Tsao DY. 2017. The code for facial identity in the primate brain. *Cell* 169:1013–28
- Chao LL, Haxby JV, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* 2:913–19
- Christensen DL, Braun KVN, Baio J, Bilder D, Charles J, et al. 2018. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveill. Summ.* 65:1–23
- Chulef AS, Read SJ, Walsh DA. 2001. A hierarchical taxonomy of human goals. *Motiv. Emot.* 25:191–232
- Cross ES, Hamilton AFdC, Grafton ST. 2006. Building a motor simulation de novo: observation of dance by dancers. *NeuroImage* 31:1257–67

- Cross ES, Hamilton AFdC, Kraemer DJM, Kelley WM, Grafton ST. 2009. Dissociable substrates for body motion and physical experience in the human action observation network. *Eur. J. Neurosci.* 30:1383–92
- Deen B, Saxe R. 2019. Parts-based representations of perceived face movements in the superior temporal sulcus. *Hum. Brain Mapp.* 2. In press. <https://doi.org/10.1002/hbm.24540>
- Digman JM. 1990. Personality structure: emergence of the five-factor model. *Annu. Rev. Psychol.* 41:417–40
- Digman JM, Takemoto-Chock NK. 1981. Factors in the natural language of personality: re-analysis, comparison, and interpretation of six major studies. *Multivar. Behav. Res.* 16:149–70
- Dinstein I, Gardner JL, Jazayeri M, Heeger DJ. 2008. Executed and observed movements have different distributed representations in human aIPS. *J. Neurosci.* 28:11231–39
- Dobs K, Bulthoff I, Schultz J. 2016. Identity information content depends on the type of facial movement. *Sci. Rep.* 6:34301
- Dobs K, Ma WJ, Reddy L. 2017. Near-optimal integration of facial form and motion. *Sci. Rep.* 7:11002
- Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. *Science* 293:2470–73
- Ekman P. 1992. Are there basic emotions? *Psychol. Rev.* 99:550–53
- Ekman P. 1999. Basic emotions. In *Handbook of Cognition and Emotion*, ed. T Dalgleish, M Power, pp. 45–60. New York: Wiley
- Ellis AW, Young AW, Critchley EMR. 1989. Loss of memory for people following temporal lobe damage. *Brain* 112:1469–83
- Ellis HD, Shepherd JW, Davies GM. 1979. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception* 8:431–39
- Engell AD, Haxby JV. 2007. Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia* 45:3234–41
- Ferrari C, Vecchi T, Todorov A, Cattaneo Z. 2016. Interfering with activity in the dorsomedial prefrontal cortex via TMS affects social impressions updating. *Cogn. Affect. Behav. Neurosci.* 16:626–34
- Fiske ST, Cuddy AJC, Glick P, Xu J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J. Personal. Soc. Psychol.* 82(6):878–902
- Fiske ST, Linville PW. 1980. What does the schema concept buy us? *Personal. Soc. Psychol. Bull.* 6:543–57
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, et al. 1995. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition* 57:109–28
- Formisano E, De Martino F, Bonte M, Goebel R. 2008. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322:970–73
- Fox CJ, Hanif HM, Iaria G, Duchaine BC, Barton JJS. 2011. Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia* 49:3188–200
- Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–51
- Frith C, Frith U. 2000. The physiological basis of theory of mind: functional neuroimaging studies. In *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, ed. S Baron-Cohen, H Tager-Flusberg, DJ Cohen, pp. 334–56. Oxford, UK: Oxford Univ. Press. 2nd ed.
- Gallagher HL, Frith CD. 2003. Functional imaging of ‘theory of mind.’ *Trends Cogn. Sci.* 7:77–83
- Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD. 2000. Reading the mind in cartoons and stories: an fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21
- Gershman SJ, Gerstenberg T, Baker CL, Cushman FA. 2016. Plans, habits, and theory of mind. *PLOS ONE* 11:e0162246
- Gershman SJ, Pouncy HT, Gweon H. 2017. Learning the structure of social influence. *Cogn. Sci.* 41:545–75
- Gobbini MI, Haxby JV. 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45:32–41
- Goodie AS, Doshi P, Young DL. 2012. Levels of theory-of-mind reasoning in competitive games. *J. Behav. Decis. Mak.* 25:95–108
- Goodman ND, Tenenbaum JB, Gerstenberg T. 2014. Concepts in a probabilistic language of thought. In *The Conceptual Mind: New Directions in the Study of Concepts*, ed. D Casasanto, G Lupyan, E Margolis, S Laurence, pp. 623–54. Cambridge, MA: MIT Press



- Grafton ST, Arbib MA, Fadiga L, Rizzolatti G. 1996. Localization of grasp representations in humans by positron emission tomography. *Exp. Brain Res.* 112:103–11
- Gray HM, Gray K, Wegner DM. 2007. Dimensions of mind perception. *Science* 315:619–19
- Grèzes J, Fonlupt P, Bertenthal B, Delon-Martin C, Segebarth C, Decety J. 2001. Does perception of biological motion rely on specific brain regions? *NeuroImage* 13:775–85
- Grill-Spector K, Weiner KS, Gomez J, Stigliani A, Natu VS. 2018. The functional neuroanatomy of face perception: from brain measurements to deep neural networks. *Interface Focus* 8:20180013
- Gruppuso V, Lindsay DS, Masson MEJ. 2007. I'd know that face anywhere! *Psychon. Bull. Rev.* 14:1085–89
- Gygax P, Oakhill J, Garnham A. 2003. The representation of characters' emotional responses: Do readers infer specific emotions? *Cogn. Emot.* 17:413–28
- Hahn CA, O'Toole AJ. 2017. Recognizing approaching walkers: neural decoding of person familiarity in cortical areas responsive to faces, bodies, and biological motion. *NeuroImage* 146:859–68
- Hahn CA, O'Toole AJ, Phillips PJ. 2016. Dissecting the time course of person recognition in natural viewing environments. *Br. J. Psychol.* 107:117–34
- Hanczakowski M, Zawadzka K, Macken B. 2015. Continued effects of context reinstatement in recognition. *Mem. Cogn.* 43:788–97
- Hasan BAS, Valdes-Sosa M, Gross J, Belin P. 2016. Hearing faces and seeing voices: amodal coding of person identity in the human brain. *Sci. Rep.* 6:37494
- Hassabis D, Spreng RN, Rusu AA, Robbins CA, Mar RA, Schacter DL. 2013. Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cereb. Cortex* 24:1979–87
- Hassin RR, Aviezer H, Bentin S. 2013. Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* 5:60–65
- Hastie R. 1980. Memory for behavioral information that confirms or contradicts a personality impression. In *Person Memory: Personality Traits as Organizing Principles in Memory for Behavior*, ed. R Hastie, TM Ostrom, EB Ebbesen, RS Wyer Jr., DL Hamilton, DE Carlson, pp. 155–78. Hillsdale, NJ: Erlbaum
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn. Sci.* 4:223–33
- Hecaen H, Angelergues R. 1962. Agnosia for faces (prosopagnosia). *Arch. Neurol.* 7:92–100
- Heider F, Simmel M. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* 57:243–59
- Hill CA, Suzuki S, Polania R, Moisa M, O'Doherty JP, Ruff CC. 2017. A causal account of the brain network computations underlying strategic social behavior. *Nat. Neurosci.* 20:1142
- Hodges JR, Patterson K, Oxbury S, Funnell E. 1992. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. *Brain* 115:1783–806
- Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. 1999. Cortical mechanisms of human imitation. *Science* 286:2526–28
- Jara-Ettinger J, Floyd S, Tenenbaum JB, Schulz LE. 2017. Children understand that agents maximize expected utilities. *J. Exp. Psychol. Gen.* 146:1574–85
- Jara-Ettinger J, Gweon H, Schulz LE, Tenenbaum JB. 2016. The naive utility calculus: computational principles underlying commonsense psychology. *Trends Cogn. Sci.* 20:589–604
- Jern A, Kemp C. 2015. A decision network account of reasoning about other people's choices. *Cognition* 142:12–38
- Jern A, Lucas CG, Kemp C. 2011. Evaluating the inverse decision-making approach to preference learning. *Adv. Neural Inform. Process. Syst.* 24:2276–84
- Jozwik K, Charest I, Kriegeskorte N, Cichy RM. 2018. *Animacy dimensions ratings and approach for decorrelating stimuli dimensions*. Paper presented at the 2018 Conference on Cognitive Computational Neuroscience, Philadelphia, PA, 5–8 Sept. <https://www2.securecms.com/CCNeuro/docs-0/59285d3868ed3f803a8a256e.pdf>
- Kaiser D, Cichy RM. 2018. Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition* 180:118–22
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17:4302–11

- Kietzmann TC, Gert AL, Tong F, Konig P. 2017. Representational dynamics of facial viewpoint encoding. *J. Cogn. Neurosci.* 29:637–51
- Konkle T, Caramazza A. 2013. Tripartite organization of the ventral stream by animacy and object size. *J. Neurosci.* 33:10235–42
- Koster-Hale J, Richardson H, Velez N, Asaba M, Young L, Saxe R. 2017. Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage* 161:9–18
- Koster-Hale J, Saxe R. 2013. Functional neuroimaging of theory of mind. In *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, ed. S Baron-Cohen, H Tager-Flusberg, DJ Cohen, pp. 132–63. Oxford, UK: Oxford Univ. Press. 3rd ed.
- Koster-Hale J, Saxe R, Dungan J, Young LL. 2013. Decoding moral judgments from neural representations of intentions. *PNAS* 110:5648–53
- Kraft-Todd GT, Reinero DA, Kelley JM, Heberlein AS, Baer L, Riess H. 2017. Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PLOS ONE* 12:e0177758
- Kriegeskorte N, Douglas PK. 2018. Cognitive computational neuroscience. *Nat. Neurosci.* 21:1148–60
- Kriegeskorte N, Formisano E, Sorger B, Goebel R. 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. *PNAS* 104:20600–5
- Kryven M. 2018. *Attributed intelligence*. PhD Diss., Waterloo Univ., Waterloo, Can.
- Kryven M, Ullman T, Cowan W, Tenenbaum J. 2016. *Outcome or strategy? A Bayesian model of intelligence attribution*. Paper presented at CogSci 2016: The Annual Meeting of the Cognitive Science Society, Philadelphia, PA
- Landi SM, Freiwald WA. 2017. Two areas for familiar face recognition in the primate brain. *Science* 357:591–95
- Leibo JZ, Liao Q, Anselmi F, Poggio T. 2015. The invariance hypothesis implies domain-specific regions in visual cortex. *PLOS Comput. Biol.* 11:e1004390
- Leshinskaya A, Contreras JM, Caramazza A, Mitchell JP. 2017. Neural representations of belief concepts: a representational similarity approach to social semantics. *Cereb. Cortex* 27:344–57
- Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM. 2000. Neural systems underlying the recognition of familiar and newly learned faces. *J. Neurosci.* 20:878–86
- Lewis MD. 2005. Bridging emotion theory and neurobiology through dynamic systems modeling. *Behav. Brain Sci.* 28:169–94
- Lingnau A, Downing PE. 2015. The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* 19:268–77
- Lowder MW, Gordon PC. 2015. Natural forces as agents: reconceptualizing the animate–inanimate distinction. *Cognition* 136:85–90
- Ma N, Vandekerckhove M, Baetens K, Van Overwalle F, Seurinck R, Fias W. 2011. Inconsistencies in spontaneous and intentional trait inferences. *Soc. Cogn. Affect. Neurosci.* 7:937–50
- Malone DR, Morris HH, Kay MC, Levin HS. 1982. Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *J. Neurol. Neurosurg. Psychiatry* 45:820–22
- Martin A, Weisberg J. 2003. Neural foundations for understanding social and mechanical concepts. *Cogn. Neuropsychol.* 20:575–87
- Mende-Siedlecki P. 2018. Changing our minds: the neural bases of dynamic impression updating. *Curr. Opin. Psychol.* 24:72–76
- Mende-Siedlecki P, Cai Y, Todorov A. 2012. The neural dynamics of updating person impressions. *Soc. Cogn. Affect. Neurosci.* 8:623–31
- Mengotti P, Dombert PL, Fink GR, Vossel S. 2017. Disruption of the right temporoparietal junction impairs probabilistic belief updating. *J. Neurosci.* 37:5419–28
- Miller GA. 1995. WordNet: a lexical database for English. *Commun. ACM* 38:39–41
- Minsky M. 2000. Commonsense-based interfaces. *Commun. ACM* 43:66–73
- Mitchell JP, Cloutier J, Banaji MR, Macrae CN. 2006. Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Soc. Cogn. Affect. Neurosci.* 1:49–55
- Mitchell JP, Macrae CN, Banaji MR. 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *J. Neurosci.* 24:4912–17
- Mitchell JP, Macrae CN, Banaji MR. 2005. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *NeuroImage* 26:251–57



- Molenberghs P, Johnson H, Henry JD, Mattingley JB. 2016. Understanding the minds of others: a neuroimaging meta-analysis. *Neurosci. Biobehav. Rev.* 65:276–91
- Molnar-Szakacs I, Kaplan J, Greenfield PM, Iacoboni M. 2006. Observing complex action sequences: the role of the fronto-parietal mirror neuron system. *NeuroImage* 33:923–35
- Natu V, Jiang F, Narvekar A, Keshvari S, O’Toole A. 2008. Representations of facial identity over changes in viewpoint. *J. Vis.* 8:159
- Negri GA, Rumiati RI, Zadini A, Ukmar M, Mahon BZ, Caramazza A. 2007. What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cogn. Neuropsychol.* 24:795–816
- Nestor A, Plaut DC, Behrmann M. 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *PNAS* 108:9998–10003
- Norman WT. 1963. Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings. *J. Abnorm. Soc. Psychol.* 66:574–83
- Ojaleto BL, Medin DL, García SG. 2017. Conceptualizing agency: folkpsychological and folkcommunicative perspectives on plants. *Cognition* 162:103–23
- Ong DC, Zaki J, Goodman ND. 2015. Affective cognition: exploring lay theories of emotion. *Cognition* 143:141–62
- Oosterhof NN, Todorov A. 2008. The functional basis of face evaluation. *PNAS* 105:11087–92
- Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE. 2010. Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *J. Neurophysiol.* 104:1077–89
- O’Toole AJ, Phillips PJ, Weimer S, Roark DA, Ayyad J, et al. 2011. Recognizing people from dynamic and static faces and bodies: dissecting identity with a fusion approach. *Vis. Res.* 51:74–83
- O’Toole AJ, Roark DA, Abdi H. 2002. Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6:261–66
- Papeo L, Wurm MF, Oosterhof NN, Caramazza A. 2017. The neural representation of human versus non-human bipeds and quadrupeds. *Sci. Rep.* 7:14040
- Peelen MV, Atkinson AP, Vuilleumier P. 2010. Supramodal representations of perceived emotions in the human brain. *J. Neurosci.* 30:10127–34
- Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G. 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb. Cortex* 15:1866–76
- Ralph MAL, Jefferies E, Patterson K, Rogers TT. 2017. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18:42–55
- Ramon M, Caharel S, Rossion B. 2011. The speed of recognition of personally familiar faces. *Perception* 40:437–49
- Rezlescu C, Barton JJS, Pitcher D, Duchaine B. 2014. Normal acquisition of expertise with greebles in two cases of acquired prosopagnosia. *PNAS* 111:5123–28
- Rizzolatti G, Craighero L. 2004. The mirror-neuron system. *Annu. Rev. Neurosci.* 27:169–92
- Rizzolatti G, Fadiga L, Matelli M, Bettinardi V, Paulesu E, et al. 1996. Localization of grasp representations in humans by PET. I. Observation versus execution. *Exp. Brain Res.* 111:246–52
- Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2:661–70
- Rizzolatti G, Sinigaglia C. 2016. The mirror mechanism: a basic principle of brain function. *Nat. Rev. Neurosci.* 17:757–65
- Russell JA. 1980. A circumplex model of affect. *J. Personal. Soc. Psychol.* 39:1161–78
- Russell JA, Bachorowski J-A, Fernández-Dols J-M. 2003. Facial and vocal expressions of emotion. *Annu. Rev. Psychol.* 54:329–49
- Russell JA, Barrett LF. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Personal. Soc. Psychol.* 76:805–19
- Saxe R, Houlihan SD. 2017. Formalizing emotion concepts within a Bayesian model of theory of mind. *Curr. Opin. Psychol.* 17:15–21

- Saxe R, Kanwisher N. 2003. People thinking about thinking people: the role of the temporo-parietal junction in theory of mind. *NeuroImage* 19:1835–42
- Schirmer A, Adolphs R. 2017. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends Cogn. Sci.* 21:216–28
- Sergent J, Ohta S, Macdonald B. 1992. Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115:15–36
- Serre T, Oliva A, Poggio T. 2007. A feedforward architecture accounts for rapid categorization. *PNAS* 104:6424–29
- Shi B, Weninger T. 2018. *Open-world knowledge graph completion*. Paper presented at the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, Feb. 2–7
- Simhi N, Yovel G. 2016. The contribution of the body and motion to whole person recognition. *Viz. Res.* 122:12–20
- Simon HA. 1978. Information-processing theory of human problem solving. In *Handbook of Learning and Cognitive Processes*, vol. 5: *Human Information Processing*, ed. WK Estes, pp. 271–95. New York: Psychology
- Skerry AE, Saxe R. 2014. A common neural code for perceived and inferred emotion. *J. Neurosci.* 34:15997–6008
- Skerry AE, Saxe R. 2015. Neural representations of emotion are organized around abstract event features. *Curr. Biol.* 25:1945–54
- Snowden JS, Thompson JC, Neary D. 2004. Knowledge of famous faces and names in semantic dementia. *Brain* 127:860–72
- Snowden JS, Thompson JC, Neary D. 2012. Famous people knowledge and the right and left temporal lobes. *Behav. Neurol.* 25:35–44
- Speer R, Chin J, Havasi C. 2017. ConceptNet 5.5: an open multilingual graph of general knowledge. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 4444–51. Menlo Park, CA: AAAI
- Spiers HJ, Love BC, Le Pelley ME, Gibb CE, Murphy RA. 2017. Anterior temporal lobe tracks the formation of prejudice. *J. Cogn. Neurosci.* 29:530–44
- Srinivasan R, Golomb JD, Martinez AM. 2016. A neural basis of facial action recognition in humans. *J. Neurosci.* 36:4434–42
- Srull TK, Wyer RS Jr. 1989. Person memory and judgment. *Psychol. Rev.* 96:58–83
- Tamir DI, Thornton MA. 2018. Modeling the predictive social mind. *Trends Cogn. Sci.* 22:201–12
- Tenenbaum JB. 1998. Mapping a manifold of perceptual observations. In *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems*, pp. 682–88. Cambridge, MA: MIT Press
- Thornton MA, Mitchell JP. 2017. Theories of person perception predict patterns of neural activity during mentalizing. *Cereb. Cortex* 28:3505–20
- Thornton MA, Tamir DI. 2017. Mental models accurately predict emotion transitions. *PNAS* 114:5982–87
- Thornton MA, Weaverdyck ME, Tamir D. 2018. The brain represents people as the mental states they habitually experience. PsyArXiv. <https://doi.org/10.31234/osf.io/uhbpz>
- Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381:520–22
- Todorov A, Uleman JS. 2002. Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. *J. Personal. Soc. Psychol.* 83:1051–65
- Todorov A, Uleman JS. 2003. The efficiency of binding spontaneous trait inferences to actors' faces. *J. Exp. Soc. Psychol.* 39:549–62
- Tomasello M, Carpenter M, Call J, Behne T, Moll H. 2005. Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28:675–91
- Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N. 2000. Response properties of the human fusiform face area. *Cogn. Neuropsychol.* 17:257–80
- Tupes EC, Christal RE. 1992. Recurrent personality factors based on trait ratings. *J. Personal.* 60:225–51
- Vannuscorps G, Caramazza A. 2016. Typical action perception and interpretation without motor simulation. *PNAS* 113:86–91
- Wagner HL, MacDonald CJ, Manstead AS. 1986. Communication of individual emotions by spontaneous facial expressions. *J. Personal. Soc. Psychol.* 50:737–43



- Wang Y, Collins JA, Koski J, Nugiel T, Metoki A, Olson IR. 2017. Dynamic neural architecture for social knowledge retrieval. *PNAS* 114:E3305–14
- Watson CE, Cardillo ER, Ianni GR, Chatterjee A. 2013. Action concepts in the brain: an activation likelihood estimation meta-analysis. *J. Cogn. Neurosci.* 25:1191–205
- Waytz A, Young L. 2018. Morality for us versus them. In *Atlas of Moral Psychology*, ed. K Gray, J Graham, pp. 186–92. New York: Guilford
- Winell M. 1987. Personal goals: the key to self-direction in adulthood. In *Humans as Self-Constructing Living Systems: Putting the Framework to Work*, ed. ME Ford, DH Ford, pp. 261–87. Hillsdale, NJ: Erlbaum
- Wu Y, Schulz LE. 2018. Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child Dev.* 89:649–62
- Wurm MF, Ariani G, Greenlee MW, Lingnau A. 2016. Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cereb. Cortex* 26:3390–401
- Wurm M, Caramazza A. 2018. Representation of action concepts in left posterior temporal cortex that generalize across vision and language. bioRxiv 361220
- Wurm MF, Caramazza A, Lingnau A. 2017. Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *J. Neurosci.* 37:562–75
- Wurm MF, Lingnau A. 2015. Decoding actions at different levels of abstraction. *J. Neurosci.* 35:7727–35
- Yee E, Jones MN, McRae K. 2018. Semantic memory. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, vol. 3: *Language and Thought*, ed. JT Wixted, S Thompson-Schill, pp. 319–56. New York: Wiley. 4th ed.
- Young AW, Newcombe F, de Haan EHF, Small M, Hay DC. 1993. Face perception after brain injury: selective impairments affecting identity and expression. *Brain* 116:941–59
- Yovel G, O'Toole AJ. 2016. Recognizing people in motion. *Trends Cogn. Sci.* 20:383–95
- Zaitchik D, Walker C, Miller S, LaViolette P, Feczko E, Dickerson BC. 2010. Mental state attribution and the temporoparietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia* 48:2528–36