

# The Acquisition of Person Knowledge

## Annual Review of Psychology

Stefano Anzellotti<sup>1</sup> and Liane L. Young<sup>1</sup>

<sup>1</sup>Department of Psychology, Boston College, Boston, USA, 02467  
stefano.anzellotti@bc.edu, liane.young@bc.edu

March 22, 2019

### **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

<b>1</b>	<b>INTRODUCTION</b>	<b>3</b>
<b>2</b>	<b>PERSON PERCEPTION</b>	<b>3</b>
2.1	Recognizing agents and conspecifics . . . . .	3
2.1.1	Recognition of agents from static images . . . . .	3
2.1.2	Motion, agency and animacy . . . . .	4
2.2	Recognizing individuals . . . . .	5
2.2.1	Familiar faces and people . . . . .	6
2.3	Recognizing facial expressions . . . . .	6
2.4	Recognizing actions . . . . .	8
2.4.1	Neural mechanisms for action recognition . . . . .	8
<b>3</b>	<b>PERSON KNOWLEDGE</b>	<b>9</b>
3.1	Emotions . . . . .	9
3.1.1	Structure . . . . .	10
3.1.2	Inference . . . . .	11
3.1.3	Neural bases . . . . .	11
3.2	Beliefs and intentions . . . . .	11
3.2.1	Structure . . . . .	11
3.2.2	Inference . . . . .	12
3.2.3	Neural bases . . . . .	12
3.3	Traits . . . . .	13
3.3.1	Structure . . . . .	13
3.3.2	Inference . . . . .	13
3.3.3	Neural bases . . . . .	14
<b>4</b>	<b>OUTSTANDING QUESTIONS</b>	<b>14</b>
4.1	How is person knowledge acquired from observation? . . . . .	15
4.2	Towards integrated person models . . . . .	15
4.3	Person model impairments . . . . .	15
<b>5</b>	<b>CONCLUSIONS</b>	<b>16</b>
<b>6</b>	<b>ELEMENTS OF THE MANUSCRIPT</b>	<b>17</b>
6.1	Figures . . . . .	17
6.2	Acronyms and abbreviations . . . . .	17
6.3	Summary Points . . . . .	18
6.4	Future issues . . . . .	18

# 1 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.

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 work in this area, bringing together the literatures on social perception and person knowledge.

## 2 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 been sometimes used in the literature to refer to a variety of processes, including some that have little to do with perception itself (i.e. 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.

### 2.1 Recognizing agents and conspecifics

Recognizing an entity as an agent and as a conspecific are 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 his/her own behavior.

#### 2.1.1 Recognition of agents from static images

Entities can be visually recognized as agents using static information such as shape as well as other cues (i.e. 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 20ms, and behavioral responses are produced as early as 290ms after stimulus onset [Thorpe et al., 1996]. The 290ms include the time to plan and execute

a motor response: EEG data indicate that differences between the neural responses to animals and inanimate objects emerge as early as 150ms post stimulus onset [Thorpe et al., 1996]. It has been argued that the speed of categorization suggests that this process is largely feedforward [Serre et al., 2007].

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 and 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 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, comprised between the object-selective regions in the medial fusiform gyrus and the dorsal stream [Konkle and 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, 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 MEG signals as early as 100ms 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 based on response patterns in ventral prefrontal cortex [Anzellotti and Caramazza, 2014a].

### 2.1.2 Motion, agency and animacy

Motion cues play a critical role for the recognition of agents. Humans attribute goals and intentions even to geometric shapes that appear to move intentionally [Heider and 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 temporal sulcus (pSTS, Pelphrey et al. [2005]). Identifying regions distinguishing humans from other animals based on 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 chicken to control for pedalism, and found that information distinguishing point-light displays of humans from animals across pedalism (that is, for both infants and adults - Papeo et al. [2017]) could be decoded in left pSTS and posterior cingulate. The right STS encodes information distinguishing bipeds from quadrupeds, but in right STS responses to 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 and Simmel, 1944] elicit activity in the pSTS bilaterally and in lateral portions of fusiform gyrus approximately corresponding to the areas which show increased responses to animals [Chao et al., 1999, Konkle and Caramazza, 2013]. This finding sug-

gests 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 solely driven 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 like wind, rain, and avalanches have agency without animacy [Lowder and 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 Ngöbe of Panama are more likely than US citizens to attribute intentions to plants and to infer that they engage in social actions like kin altruism [Medin et al., 2017].

## 2.2 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 and 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 only generalize across changes in viewpoint [Anzellotti et al., 2013, Anzellotti and Caramazza, 2014b], while the anterior temporal lobe also generalizes across face parts [Anzellotti and Caramazza, 2015].<sup>1</sup> A similar organization consisting of selective patches with different degrees of generalization is observed in macaque monkeys [Freiwald and 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 and 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 weighting different cues depending on their reliability [Dobs et al., 2017]. The identity of moving individuals can be decoded from STS, and, when the individuals are close to the observer, from body-selective regions like the extrastriate body area and the fusiform body area [Hahn et al., 2016, Hahn and O’toole, 2017]. Voice identity can also be decoded from the STS [Formisano et al., 2008, Anzellotti and Caramazza, 2017, Hasan et al., 2016]. Furthermore, STS encodes representations of person identity that generalize across the visual and auditory modality [Anzellotti and Caramazza, 2017].

---

<sup>1</sup>Different temporal stages of face processing have been recently identified with electroencephalography [Kietzmann et al., 2017].

In addition to 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 and Cichy, 2018, Castelhana and Pereira, 2017]. 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 been previously shown increases false alarm rates for judgments of face familiarity [Gruppuso et al., 2007].

### 2.2.1 Familiar faces and people

Familiar faces are recognized more accurately than unfamiliar faces using part of the face, and are recognized more accurately given presentation of the inner part of the face than given 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 the faces are obscured. This suggests that information about the face contributes importantly, and person recognition could not be completed at the same level of accuracy relying only on the body. Taken together, this evidence shows 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 unfamiliar faces in several cortical regions, including posterior cingulate, medial prefrontal cortex, anterior STS, and hippocampus [Leveroni et al., 2000, Gobbini and 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 medial parietal regions presented with 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 medial parietal cortex (i.e. posterior cingulate) for 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 anterior temporal lobe (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 one in entorhinal cortex [Landi and Freiwald, 2017]. The relative contributions of posterior cingulate and ATL remain unknown.

## 2.3 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 and Russell, 1996, Barrett et al., 2011, Hassin et al., 2013].

However, spontaneous facial expressions still convey sufficient information to infer emotions with above-chance accuracy<sup>2</sup> [Wagner et al., 1986], and expressions can be used to infer desires and beliefs [Wu and Schulz, 2018]. Furthermore, facial expressions might be used to disambiguate between alternative possible emotional reactions to a given context [Saxe and 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]: ventral temporal regions (OFA and FFA) would be specialized for face identity, while pSTS would be specialized for facial expressions. Indeed, in pSTS, emotional faces yield a stronger response than neutral faces [Engell and Haxby, 2007], and produce response patterns that can be used to decode emotional valence [Skerry and Saxe, 2014]. Multivariate analyses show that the pSTS encodes information about emotion that generalizes across facial expressions, voices, body posture [Peelen et al., 2010, Skerry and Saxe, 2014], and information about specific face movements [Srinivasan et al., 2016, Deen and Saxe, 2019].

However, support for a complete separation between identity recognition and expression recognition is weaker than is often assumed [Calder and Young, 2005, Bernstein and Yovel, 2015]. Emotional faces yield stronger responses than neutral faces not only in pSTS but also in occipital and fusiform regions [Engell and Haxby, 2007], and the valence of expressions can be decoded from the OFA and the FFA [Skerry and Saxe, 2014]. In addition, recent results show that face identity can be decoded from patterns of activity in pSTS [Anzellotti and Caramazza, 2017], and a patient with a lesion involving the pSTS (Fox et al. [2011], patient 5) showed impairments not only for the recognition of facial expressions, but also for the recognition of identity across different expressions.

One possible explanation for the finding that identity can be decoded from 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 to isolate what aspects of an image are due to expression, and vice versa.<sup>3</sup>

If recognition of identity and 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 [O’Neill et al., 2019].

---

<sup>2</sup>Here a participant’s response is considered as ‘accurate’ if it matches 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.

<sup>3</sup>This hypothesis is consistent with the proposal that pSTS might also contribute to recognition of identity from dynamic stimuli [O’Toole et al., 2002].

The broader area surrounding the face-selective pSTS might play a more general role for 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 and Yovel, 2016]. Regions of pSTS neighboring the face-selective pSTS respond to biological motion and to point light displays [Martin and Weisberg, 2003]. Taken together, this evidence suggests that a larger patch of pSTS including the face-selective pSTS might be a multimodal convergence zone integrating motion and form information as well as auditory information [Yovel and O’Toole, 2016, Peelen et al., 2010, Anzellotti and Caramazza, 2017].

## 2.4 Recognizing actions

Recognizing the actions of other agents is critical for detecting threats, engaging in cooperation, and coordinating our own actions with consideration to 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 based on the category of the object involved: we can recognize the opening or closing of a specific bottle, of any bottle, or of any object (i.e. a bottle and a box, Wurm and Lingnau [2015]). Actions can also be categorized at different levels of abstraction based on their goal. We can recognize the action of clapping, or at a more abstract level the action of producing sound.

### 2.4.1 Neural mechanisms for action recognition

Neuroimaging studies show that action observation leads to increased responses in lateral occipito-temporal cortex (LOTc) [Watson et al., 2013, Lingnau and Downing, 2015], as well as pSTS, anterior intraparietal sulcus/inferior parietal lobule (aIPS/IPL), ventral premotor cortex (PMv), and the supplementary motor area (SMA) [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’ (AON) [Calvo-Merino et al., 2006, Cross et al., 2009]. A similar network of brain regions is activated by biological motion [Grezes et al., 2001].

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



and the aIPS/IPL representations of actions [Wurm and Lingnau, 2015, Wurm et al., 2016] overlap.

Recent work shows that LOTC is organized at a macroscopic scale by the transitivity and sociality of actions. Pattern similarity in dorsal portions of 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 and Craighero, 2004, Rizzolatti and Sinigaglia, 2016]: action understanding consists in a ‘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 be used just as well 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 and Lingnau, 2015, Wurm et al., 2017], which does not respond during action execution. In addition, patients with impairments for action execution [Negri et al., 2007] and patients with upper limb dysplasia [Vannuscorps and Caramazza, 2016] can have spared action recognition.

### 3 PERSON KNOWLEDGE

Humans represent a wealth of information about others, ranging from someone’s current mental states (i.e. emotions and thoughts), to more lasting traits (i.e. personality and moral values), from semantic knowledge (i.e. someone’s occupation) to episodic memories of particular moment spent with someone (i.e. meeting family arriving at the airport). Understanding person knowledge requires understanding 1) what information we represent about others, 2) how this information is acquired, and 3) how this information is used to make new inferences and decisions. We will discuss in separate subsections emotions, beliefs and intentions, and traits, but this does not amount to a claim that they are distinct natural kinds. Due to space limitations, for semantic knowledge we refer the reader to recent reviews [Ralph et al., 2017, Leshinskaya et al., 2017, Yee et al., 2018].

#### 3.1 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 will not be discussed in detail (i.e. Barrett [2014]).

### 3.1.1 Structure

The theory of basic emotions [Ekman, 1992] proposes that emotions can be represented as a vectors of 5 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, Feldman Barrett and Russell, 1998, Russell and 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 not be best represented as a vector space, but as a manifold (i.e. see Tenenbaum [1998]).

Recent models suggest that more than 5 dimensions might be needed to capture human emotion attribution. A space consisting of 38 appraisal dimensions was found to outperform a model using 5 basic emotions and a model using valence and arousal [Skerry and Saxe, 2015]. An optimized 10 dimensional space could achieve very similar performance to the 38 appraisal dimensions [Skerry and 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 and Houlihan, 2017]. The language of probability also helps to differentiate between 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 sum, representations of emotions could be thought of as consisting of an emotion space or manifold, equipped with 1) a similarity metric and 2) a stochastic process that captures the non-independence between emotion dimensions and the dependence of emotions on their history (see Lewis [2005], Thornton and Tamir [2017], Tamir and Thornton [2018]). Individual-specific stochastic processes can be learned, representing information about the temporal dynamics of an emotion for that individual (i.e. ‘does he hold a grudge?’), and the interactions between different emotions (i.e. ‘does she get upset when she is surprised?’). These issues bring us closer to the topic of traits which will be discussed later.

### 3.1.2 Inference

Observers can infer emotions from several types of cues, including facial expressions, actions, and situations [Gygax et al., 2003, Skerry and Saxe, 2014]. Recent studies propose to understand emotion attribution using ideas from probability theory and Bayesian model inversion [Ong et al., 2015, Saxe and 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 and Tenenbaum, 2014, Baker et al., 2017], intentions [Jern and 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 (i.e. a situation) lead to non-observable emotional states, which in turn lead to observable actions and facial expressions. As a consequence, the probability of non-observable 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 likely resulting emotional states and the inversion of a model that links the emotional states to the observed actions and facial expressions [Saxe and Houlihan, 2017].

### 3.1.3 Neural bases

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

## 3.2 Beliefs and intentions

### 3.2.1 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<sup>4</sup>. 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 and 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.

---

<sup>4</sup>Although of course the appropriate theories for the representation of one’s own beliefs and for the representation of others’ beliefs might be different.

Several models have been proposed for the structure of intentions/goals [Austin and Vancouver, 1996]. The most influential accounts are based on a hierarchy of goals organized in clusters at different levels (i.e. Chulef et al. [2001]). Dimensional accounts of goals have also been proposed (i.e. Winell [1987]), leading to 6 proposed factors: ‘importance’, ‘difficulty’, ‘specificity’, ‘temporal range’, ‘level of consciousness’, and ‘connectedness’ [Austin and Vancouver, 1996].

### 3.2.2 Inference

A wealth of research has investigated the development of intention understanding, for lack of space we refer the reader to existing reviews [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 and Tenenbaum, 2014, Baker et al., 2017]. The framework of Partially Observable Markov Decision Processes (POMDPs, Cassandra [1998]) has been used to account for how multiple sequential observations are integrated during inference [Baker and Tenenbaum, 2014, Baker et al., 2017].

### 3.2.3 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 sub-regions of the medial prefrontal cortex (DMPFC, VMPFC), right and left temporo-parietal junction (RTPJ, LTPJ), and precuneus. Brain regions in the ‘Theory of Mind Network’ or ToM [Frith and Frith, 2000, Gallagher and 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 like melting or rusting and representing ‘false’ photographs or maps [Saxe and Kanwisher, 2003], and more to sentences describing thoughts than facts [Zaitchik et al., 2010]. A recent activation likelihood estimation (ALE) meta-analysis of 144 datasets (3150 participants) uncovered MPFC and bilateral TPJ activation across all ToM tasks sampled [Molenberghs et al., 2016]. In sum, regions in the ToM network respond during the attribution of beliefs to others across a variety of experimental paradigms (see Koster-Hale and Saxe [2013] for an in-depth review).

Patterns of activity in 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 TMS evidence that RTPJ contributes to updating probabilistic beliefs). Other recent work has combined inhibitory continuous theta-burst TMS with model-based 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 player’s strategy, as well as the functional connectivity of the RTPJ to VMPFC and DMPFC.

### 3.3 Traits

In addition to mental states, beliefs, and intentions, observers attribute to others more lasting properties - traits [Allport and Odbert, 1936]. Whether or not states and traits are qualitatively different is an open issue in the literature [Allen and Potkay, 1981, Anzellotti, 2019].

#### 3.3.1 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 and Christal, 1992, Norman, 1963], and across very different populations of participants, such as teachers rating children and college students rating one another [Digman and Takemoto-Chock, 1981].

Other dimensions capturing traits have been proposed in the literature, such as warmth and competence [Fiske et al., 2018], agency and experience [Gray et al., 2007], and trustworthiness and social dominance [Oosterhof and 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 and 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 and Mitchell, 2017].

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 ‘sums of states’ model has been shown to outperform the previous three dimensional model [Thornton and Mitchell, 2017] at explaining neural responses [Thornton et al., 2018]. However, the sums of states model only accounts 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.

#### 3.3.2 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 and 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

and 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 vs optimal strategies. This work takes initial steps towards the long-term goal of building task-performing models that can produce human-like trait inferences based on perceptual inputs.

### 3.3.3 Neural bases

In the section on perception of familiar people, we discussed the involvement of posterior cingulate, anterior STS and hippocampus, and medial prefrontal cortex (mPFC). Early social neuroscience studies found stronger responses in 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, 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, mPFC responds more to action that are diagnostic of traits than to actions that are not (i.e. ‘he played his music loud at the public picnic grounds’ vs ‘he ordered a cup of coffee at Starbucks’) [Mitchell et al., 2006]. Response patterns in mPFC also distinguish between individuals with high and low agreeableness, and between individuals with different trait combinations [Hassabis et al., 2013].

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

In addition to 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 mPFC [Hassabis et al., 2013]. Furthermore, a recent study found that the anterior temporal lobe (ATL) tracks the valence associated with social groups [Spiers et al., 2017]. A wealth of neuropsychological studies implicates ATL in the representation of semantic knowledge about people [Ellis et al., 1989, Snowden et al., 2012] - its role for the representation of traits remains to be elucidated.

## 4 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.

#### 4.1 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 and Adolphs, 2017, Grill-Spector et al., 2018]. Second, it requires a consideration of the integration of observations about an agent and about context [Carroll and Russell, 1996, Aviezer et al., 2012, Saxe and Houlihan, 2017, Baker et al., 2017]. Third, it requires us to characterize the acquisition of information over time, and 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.

#### 4.2 Towards integrated person models

Most studies in the literature adopt a ‘divide et impera’ 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 and 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, O’Neill et al., 2019], of goals and beliefs [Baker and Tenenbaum, 2014], of states and traits [Tamir and Thornton, 2018, Anzellotti, 2019], of group membership and moral judgments [Waytz and 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 does not mean just understanding 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.

#### 4.3 Person model impairments

Finally, disorders of social cognition such as autism spectrum disorder 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 help to pinpoint quantitatively which components of the computations are affected in patients, and lead to a clearer picture of the underlying neural impairments.

## 5 CONCLUSIONS

We have attempted to collect in one article 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 and Douglas [2018] for broader discussion of task-performing models). These models will need to infer unobservable latent variables like 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 (for example in the moral domain). Moving towards this goal will call for growing 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.



## 6 ELEMENTS OF THE MANUSCRIPT

### 6.1 Figures

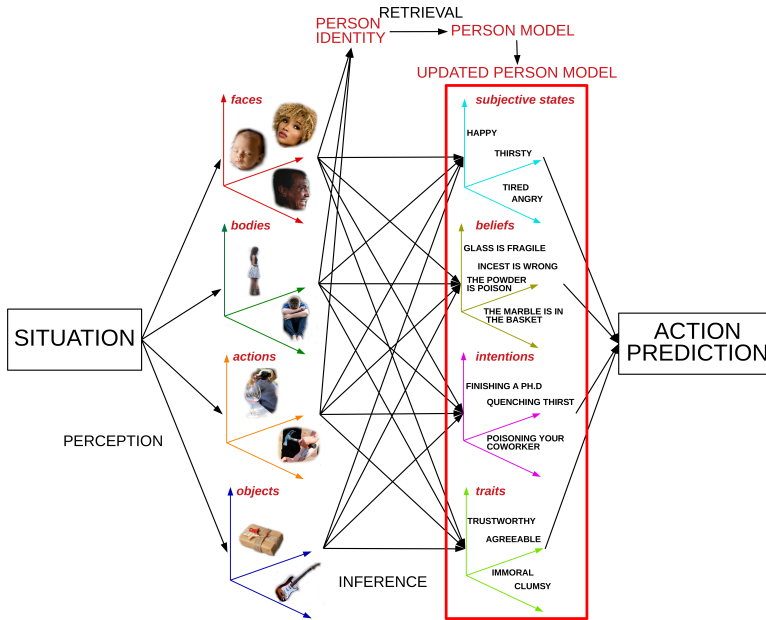


Figure 1: Schematic of the mechanisms engaged in the understanding and prediction of other people. Sensory inputs produced by a situation involving a person are mapped onto perceptual representations. Recognition of the identity of the person is used to retrieve the prior state of the person model for 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 only meant 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.

### 6.2 Acronyms and abbreviations

1. fMRI: functional Magnetic Resonance Imaging
2. TMS: transcranial magnetic stimulation
3. ERP: event related potentials
4. ToM: theory of mind
5. ASD: autism spectrum disorders
6. FFA: fusiform face area
7. STS: superior temporal sulcus

8. TPJ: temporoparietal junction
9. ATL: anterior temporal lobe
10. MPFC: medial prefrontal cortex

### **6.3 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, ATL for faces, EBA, FBA for bodies, LOTC, aIPS, PMv for actions).
6. ‘Theory of Mind’ regions, including the temporoparietal junction (TPJ), medial prefrontal cortex (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.

### **6.4 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 would like to 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

### References

- Bem P Allen and Charles R Potkay. On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology*, 41(5):916, 1981.
- Gordon W Allport and Henry S Odbert. Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1):i, 1936.
- John R Anderson. Concepts, propositions, and schemata: What are the cognitive units? Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF PSYCHOLOGY, 1980.
- Stefano Anzellotti. Modeling the relation between states and traits. 2019.
- Stefano Anzellotti and Alfonso Caramazza. Individuating the neural bases for the recognition of conspecifics with mvpa. *Neuroimage*, 89:165–170, 2014a.
- Stefano Anzellotti and Alfonso Caramazza. The neural mechanisms for the recognition of face identity in humans. *Frontiers in psychology*, 5:672, 2014b.
- Stefano Anzellotti and Alfonso Caramazza. From parts to identity: invariance and sensitivity of face representations to different face halves. *Cerebral Cortex*, 26(5):1900–1909, 2015.
- Stefano Anzellotti and Alfonso Caramazza. Multimodal representations of person identity individuated with fmri. *Cortex*, 89:85–97, 2017.
- Stefano Anzellotti, Scott L Fairhall, and Alfonso Caramazza. Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, 24(8):1988–1995, 2013.
- James T Austin and Jeffrey B Vancouver. Goal constructs in psychology: Structure, process, and content. *Psychological bulletin*, 120(3):338, 1996.
- Hillel Aviezer, Yaacov Trope, and Alexander Todorov. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338(6111):1225–1229, 2012.
- Chris L Baker and Joshua B Tenenbaum. Modeling human plan recognition using bayesian theory of mind. *Plan, activity, and intent recognition: Theory and practice*, pages 177–204, 2014.
- Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.
- Chris L Baker, Julian Jara-Ettinger, Rebecca Saxe, and Joshua B Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4):0064, 2017.
- Moshe Bar. Visual objects in context. *Nature Reviews Neuroscience*, 5(8):617, 2004.

- Moshe Bar, Karim S Kassam, Avniel Singh Ghuman, Jasmine Boshyan, Annette M Schmid, Anders M Dale, Matti S Hamalainen, Ksenija Marinkovic, Daniel L Schacter, Bruce R Rosen, et al. Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*, 103(2):449–454, 2006.
- Sean G Baron, MI Gobbi, Andrew D Engell, and Alexander Todorov. Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, 6(5):572–581, 2010.
- Lisa Feldman Barrett. The conceptual act theory: A precis. *Emotion Review*, 6(4):292–297, 2014.
- Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. Context in emotion perception. *Current Directions in Psychological Science*, 20(5):286–290, 2011.
- Jean Yves Baudouin, Stephane Sansone, and Guy Tiberghien. Recognizing expression from familiar and unfamiliar faces. *Pragmatics & Cognition*, 8(1):123–146, 2000.
- Michal Bernstein and Galit Yovel. Two neural pathways of face processing: a critical evaluation of current models. *Neuroscience & Biobehavioral Reviews*, 55:536–546, 2015.
- Randolph Blake, Lauren M Turner, Moria J Smoski, Stacie L Pozdol, and Wendy L Stone. Visual recognition of biological motion is impaired in children with autism. *Psychological science*, 14(2):151–157, 2003.
- Giovanni Buccino, Ferdinand Binkofski, Gereon R Fink, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, Rudiger J Seitz, Karl Zilles, Giacomo Rizzolatti, and H-J Freund. Action observation activates premotor and parietal areas in a somatotopic manner: an fmri study. *European journal of neuroscience*, 13(2):400–404, 2001.
- A Mike Burton, Stephen Wilson, Michelle Cowan, and Vicki Bruce. Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10(3):243–248, 1999.
- Andrew J Calder and Andrew W Young. Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8):641, 2005.
- Beatriz Calvo-Merino, Julie Grezes, Daniel E Glaser, Richard E Passingham, and Patrick Haggard. Seeing or doing? influence of visual and motor familiarity in action observation. *Current biology*, 16(19):1905–1910, 2006.
- Alfonso Caramazza and Bradford Z Mahon. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in cognitive sciences*, 7(8):354–361, 2003.
- Alfonso Caramazza, Stefano Anzellotti, Lukas Strnad, and Angelika Lingnau. Embodied cognition and mirror neurons: a critical assessment. *Annual review of neuroscience*, 37:1–15, 2014.

- James M Carroll and James A Russell. Do facial expressions signal specific emotions? judging emotion from the face in context. *Journal of personality and social psychology*, 70(2):205, 1996.
- Anthony R Cassandra. A survey of pomdp applications. 1724, 1998.
- Monica S Castelhana and Effie J Pereira. The influence of scene context on parafoveal processing of objects. *The Quarterly Journal of Experimental Psychology*, (just-accepted):1–42, 2017.
- Maxime Cauchoix, Gladys Barragan-Jason, Thomas Serre, and Emmanuel J Barbeau. The neural dynamics of face detection in the wild revealed by mvpa. *Journal of Neuroscience*, 34(3):846–854, 2014.
- Le Chang and Doris Y Tsao. The code for facial identity in the primate brain. *Cell*, 169(6):1013–1028, 2017.
- Linda L Chao, James V Haxby, and Alex Martin. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, 2(10):913, 1999.
- Deborah L Christensen, Kim Van Naarden Braun, Jon Baio, Deborah Bilder, Jane Charles, John N Constantino, Julie Daniels, Maureen S Durkin, Robert T Fitzgerald, Margaret Kurzius-Spencer, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2012. *MMWR Surveillance Summaries*, 65(13):1, 2018.
- Ada S Chulef, Stephen J Read, and David A Walsh. A hierarchical taxonomy of human goals. *Motivation and Emotion*, 25(3):191–232, 2001.
- Emily S Cross, Antonia F de C Hamilton, and Scott T Grafton. Building a motor simulation de novo: observation of dance by dancers. *Neuroimage*, 31(3):1257–1267, 2006.
- Emily S Cross, Antonia F de C Hamilton, David JM Kraemer, William M Kelley, and Scott T Grafton. Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, 30(7):1383–1392, 2009.
- Ben Deen and Rebecca Saxe. Parts-based representations of perceived face movements in the superior temporal sulcus. *Human Brain Mapping*, 2 2019. ISSN 1065-9471. doi: 10.1002/hbm.24540. URL <https://doi.org/10.1002/hbm.24540>.
- John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.
- John M Digman and Naomi K Takemoto-Chock. Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate behavioral research*, 16(2):149–170, 1981.
- Ilan Dinstein, Justin L Gardner, Mehrdad Jazayeri, and David J Heeger. Executed and observed movements have different distributed representations in human aips. *Journal of Neuroscience*, 28(44):11231–11239, 2008.

- Katharina Dobs, Isabelle Bulthoff, and Johannes Schultz. Identity information content depends on the type of facial movement. *Scientific reports*, 6:34301, 2016.
- Katharina Dobs, Wei Ji Ma, and Leila Reddy. Near-optimal integration of facial form and motion. *Scientific reports*, 7(1):11002, 2017.
- Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.
- Paul Ekman. Are there basic emotions? 1992.
- Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, pages 45–60, 1999.
- Andrew W Ellis, Andrew W Young, and Edmund MR Critchley. Loss of memory for people following temporal lobe damage. *Brain*, 112(6):1469–1483, 1989.
- Hadyn D Ellis, John W Shepherd, and Graham M Davies. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8(4):431–439, 1979.
- Andrew D Engell and James V Haxby. Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, 45(14):3234–3241, 2007.
- Lisa Feldman Barrett and James A Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4):967, 1998.
- Chiara Ferrari, Tomaso Vecchi, Alexander Todorov, and Zaira Cattaneo. Interfering with activity in the dorsomedial prefrontal cortex via tms affects social impressions updating. *Cognitive, Affective, & Behavioral Neuroscience*, 16(4):626–634, 2016.
- Susan T Fiske and Patricia W Linville. What does the schema concept buy us? *Personality and social psychology bulletin*, 6(4):543–557, 1980.
- Susan T Fiske, Amy JC Cuddy, Peter Glick, and Jun Xu. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition (2002). In *Social Cognition*, pages 171–222. Routledge, 2018.
- Paul C Fletcher, Francesca Happe, Uta Frith, Simon C Baker, Ray J Dolan, Richard SJ Frackowiak, and Chris D Frith. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition*, 57(2):109–128, 1995.
- Elia Formisano, Federico De Martino, Milene Bonte, and Rainer Goebel. ” who” is saying” what”? brain-based decoding of human voice and speech. *Science*, 322(5903):970–973, 2008.
- Christopher J Fox, Hashim M Hanif, Giuseppe Iaria, Bradley C Duchaine, and Jason JS Barton. Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, 49(12):3188–3200, 2011.

- Winrich A Freiwald and Doris Y Tsao. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005):845–851, 2010.
- Chris Frith and Uta Frith. The physiological basis of theory of mind: functional neuroimaging studies. *Understanding other minds: Perspectives from developmental cognitive neuroscience*, 2, 2000.
- Helen L Gallagher and Christopher D Frith. Functional imaging of ‘theory of mind’. *Trends in cognitive sciences*, 7(2):77–83, 2003.
- Helen L Gallagher, Francesca Happe, Nicola Brunswick, Paul C Fletcher, Uta Frith, and Chris D Frith. Reading the mind in cartoons and stories: an fmri study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia*, 38(1):11–21, 2000.
- Samuel J Gershman, Tobias Gerstenberg, Chris L Baker, and Fiery A Cushman. Plans, habits, and theory of mind. *PloS one*, 11(9):e0162246, 2016.
- Samuel J Gershman, Hillard Thomas Pouncy, and Hyowon Gweon. Learning the structure of social influence. *Cognitive science*, 41:545–575, 2017.
- M Ida Gobbini and James V Haxby. Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1):32–41, 2007.
- Adam S Goodie, Prashant Doshi, and Diana L Young. Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, 25(1):95–108, 2012.
- Noah D Goodman, Joshua B Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. 2014.
- Scott T Grafton, Michael A Arbib, Luciano Fadiga, and Giacomo Rizzolatti. Localization of grasp representations in humans by positron emission tomography. *Experimental brain research*, 112(1):103–111, 1996.
- Heather M Gray, Kurt Gray, and Daniel M Wegner. Dimensions of mind perception. *science*, 315(5812):619–619, 2007.
- Julie Grezes, Pierre Fonlupt, Bennett Bertenthal, Chantal Delon-Martin, Christoph Segebarth, and Jean Decety. Does perception of biological motion rely on specific brain regions? *Neuroimage*, 13(5):775–785, 2001.
- Kalanit Grill-Spector, Kevin S Weiner, Jesse Gomez, Anthony Stigliani, and Vaidehi S Natu. The functional neuroanatomy of face perception: from brain measurements to deep neural networks. *Interface Focus*, 8(4):20180013, 2018.
- Vincenza Gruppuso, D Stephen Lindsay, and Michael EJ Masson. I’d know that face anywhere! *Psychonomic Bulletin & Review*, 14(6):1085–1089, 2007.
- Pascal Gygas, Jane Oakhill, and Alan Garnham. The representation of characters’ emotional responses: Do readers infer specific emotions? *Cognition and Emotion*, 17(3):413–428, 2003.



- Carina A Hahn and Alice J O’toole. Recognizing approaching walkers: Neural decoding of person familiarity in cortical areas responsive to faces, bodies, and biological motion. *NeuroImage*, 146:859–868, 2017.
- Carina A Hahn, Alice J O’toole, and P Jonathon Phillips. Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*, 107(1):117–134, 2016.
- Maciej Hanczakowski, Katarzyna Zawadzka, and Bill Macken. Continued effects of context reinstatement in recognition. *Memory & Cognition*, 43(5):788–797, 2015.
- Bashar Awwad Shiekh Hasan, Mitchell Valdes-Sosa, Joachim Gross, and Pascal Belin. “hearing faces and seeing voices”: Amodal coding of person identity in the human brain. *Scientific reports*, 6:37494, 2016.
- Demis Hassabis, R Nathan Spreng, Andrei A Rusu, Clifford A Robbins, Raymond A Mar, and Daniel L Schacter. Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cerebral Cortex*, 24(8):1979–1987, 2013.
- Ran R Hassin, Hillel Aviezer, and Shlomo Bentin. Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, 5(1):60–65, 2013.
- Reid Hastie. Memory for behavioral information that confirms or contradicts a personality impression. *Hastie, R.; Ostrom, TM.; Ebbesen, EB*, pages 155–178, 1980.
- James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233, 2000.
- H Hecaen and R Angelergues. Agnosia for faces (prosopagnosia). *Archives of neurology*, 7(2):92–100, 1962.
- Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259, 1944.
- Christopher A Hill, Shinsuke Suzuki, Rafael Polania, Marius Moisa, John P O’Doherty, and Christian C Ruff. A causal account of the brain network computations underlying strategic social behavior. *Nature neuroscience*, 20(8):1142, 2017.
- John R Hodges, Karalyn Patterson, Susan Oxbury, and Elaine Funnell. Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115(6):1783–1806, 1992.
- Marco Iacoboni, Roger P Woods, Marcel Brass, Harold Bekkering, John C Mazziotta, and Giacomo Rizzolatti. Cortical mechanisms of human imitation. *science*, 286(5449):2526–2528, 1999.
- Julian Jara-Ettinger, Hyowon Gweon, Laura E Schulz, and Joshua B Tenenbaum. The naive utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8):589–604, 2016.

- Julian Jara-Ettinger, Sammy Floyd, Joshua B Tenenbaum, and Laura E Schulz. Children understand that agents maximize expected utilities. *Journal of Experimental Psychology: General*, 146(11):1574, 2017.
- Alan Jern and Charles Kemp. A decision network account of reasoning about other people’s choices. *Cognition*, 142:12–38, 2015.
- Alan Jern, Christopher G Lucas, and Charles Kemp. Evaluating the inverse decision-making approach to preference learning. pages 2276–2284, 2011.
- Kamila Jozwik, Ian Charest, Nikolaus Kriegeskorte, and Radoslaw Martin Cichy. Animacy dimensions ratings and approach for decorrelating stimuli dimensions. 2018.
- Daniel Kaiser and Radoslaw M Cichy. Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition*, 180:118–122, 2018.
- Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311, 1997.
- Tim C Kietzmann, Anna L Gert, Frank Tong, and Peter Konig. Representational dynamics of facial viewpoint encoding. *Journal of cognitive neuroscience*, 29(4):637–651, 2017.
- Talia Konkle and Alfonso Caramazza. Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, 33(25):10235–10242, 2013.
- Jorie Koster-Hale and Rebecca Saxe. Functional neuroimaging of theory of mind. *Understanding other minds: Perspectives from developmental social neuroscience*, pages 132–163, 2013.
- Jorie Koster-Hale, Rebecca Saxe, James Dungan, and Liane L Young. Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences*, 110(14):5648–5653, 2013.
- Jorie Koster-Hale, Hilary Richardson, Natalia Velez, Mika Asaba, Liane Young, and Rebecca Saxe. Mentalizing regions represent distributed, continuous, and abstract dimensions of others’ beliefs. *Neuroimage*, 161:9–18, 2017.
- Gordon T Kraft-Todd, Diego A Reiner, John M Kelley, Andrea S Heberlein, Lee Baer, and Helen Riess. Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PloS one*, 12(5):e0177758, 2017.
- Nikolaus Kriegeskorte and Pamela K Douglas. Cognitive computational neuroscience. *Nature neuroscience*, page 1, 2018.
- Nikolaus Kriegeskorte, Elia Formisano, Bettina Sorger, and Rainer Goebel. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(51):20600–20605, 2007.
- Marta Kryven. Attributed intelligence. 2018.

- Marta Kryven, Tomer Ullman, William Cowan, and Josh Tenenbaum. Outcome or strategy? a bayesian model of intelligence attribution. 2016.
- Sofia M Landi and Winrich A Freiwald. Two areas for familiar face recognition in the primate brain. *Science*, 357(6351):591–595, 2017.
- Joel Z Leibo, Qianli Liao, Fabio Anselmi, and Tomaso Poggio. The invariance hypothesis implies domain-specific regions in visual cortex. *PLoS computational biology*, 11(10):e1004390, 2015.
- Anna Leshinskaya, Juan Manuel Contreras, Alfonso Caramazza, and Jason P Mitchell. Neural representations of belief concepts: a representational similarity approach to social semantics. *Cerebral Cortex*, 27(1):344–357, 2017.
- Catherine L Leveroni, Michael Seidenberg, Andrew R Mayer, Larissa A Mead, Jeffrey R Binder, and Stephen M Rao. Neural systems underlying the recognition of familiar and newly learned faces. *Journal of Neuroscience*, 20(2):878–886, 2000.
- Marc D Lewis. Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and brain sciences*, 28(2):169–194, 2005.
- Angelika Lingnau and Paul E Downing. The lateral occipitotemporal cortex in action. *Trends in cognitive sciences*, 19(5):268–277, 2015.
- Matthew W Lowder and Peter C Gordon. Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, 136:85–90, 2015.
- Ning Ma, Marie Vandekerckhove, Kris Baetens, Frank Van Overwalle, Ruth Seurinck, and Wim Fias. Inconsistencies in spontaneous and intentional trait inferences. *Social cognitive and affective neuroscience*, 7(8):937–950, 2011.
- Daniel R Malone, Harold H Morris, Marilyn C Kay, and Harvey S Levin. Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery & Psychiatry*, 45(9):820–822, 1982.
- Alex Martin and Jill Weisberg. Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, 20(3-6):575–587, 2003.
- Douglas L Medin, Salino G Garcia, et al. Conceptualizing agency: Folkpsychological and folkcommunicative perspectives on plants. *Cognition*, 162:103–123, 2017.
- Peter Mende-Siedlecki. Changing our minds: the neural bases of dynamic impression updating. *Current opinion in psychology*, 2018.
- Peter Mende-Siedlecki, Yang Cai, and Alexander Todorov. The neural dynamics of updating person impressions. *Social cognitive and affective neuroscience*, 8(6):623–631, 2012.
- Paola Mengotti, Pascasie L Dombert, Gereon R Fink, and Simone Vessel. Disruption of the right temporoparietal junction impairs probabilistic belief updating. *Journal of Neuroscience*, 37(22):5419–5428, 2017.

- George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- Marvin Minsky. Commonsense-based interfaces. *Communications of the ACM*, 43(8):66–73, 2000.
- Jason P Mitchell, C Neil Macrae, and Mahzarin R Banaji. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, 24(21):4912–4917, 2004.
- Jason P Mitchell, C Neil Macrae, and Mahzarin R Banaji. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*, 26(1):251–257, 2005.
- Jason P Mitchell, Jasmin Cloutier, Mahzarin R Banaji, and C Neil Macrae. Medial prefrontal dissociations during processing of trait diagnostic and non-diagnostic person information. *Social Cognitive and Affective Neuroscience*, 1(1):49–55, 2006.
- Pascal Molenberghs, Halle Johnson, Julie D Henry, and Jason B Mattingley. Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 65:276–291, 2016.
- Istvan Molnar-Szakacs, Jonas Kaplan, Patricia M Greenfield, and Marco Iacoboni. Observing complex action sequences: the role of the fronto-parietal mirror neuron system. *Neuroimage*, 33(3):923–935, 2006.
- Vaidehi Natu, Fang Jiang, Abhijit Narvekar, Shaiyan Keshvari, and Alice O’Toole. Representations of facial identity over changes in viewpoint. *Journal of Vision*, 8(6):159–159a, 2008.
- Gioia AL Negri, Raffaella I Rumiati, Antonietta Zadini, Maja Ukmar, Bradford Z Mahon, and Alfonso Caramazza. What is the role of motor simulation in action and object recognition? evidence from apraxia. *Cognitive Neuropsychology*, 24(8):795–816, 2007.
- Adrian Nestor, David C Plaut, and Marlene Behrmann. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, 108(24):9998–10003, 2011.
- Warren T Norman. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66(6):574, 1963.
- Kathryn C. O’Neill, Rebecca R Saxe, and Stefano Anzellotti. Recognition of identity and expression as integrated processes. 2019.
- Desmond C Ong, Jamil Zaki, and Noah D Goodman. Affective cognition: Exploring lay theories of emotion. *Cognition*, 143:141–162, 2015.
- Nikolaas N Oosterhof and Alexander Todorov. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32):11087–11092, 2008.

- Nikolaas N Oosterhof, Alison J Wiggett, Joern Diedrichsen, Steven P Tipper, and Paul E Downing. Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, 104(2):1077–1089, 2010.
- Alice J O’Toole, Dana A Roark, and Herve Abdi. Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences*, 6(6):261–266, 2002.
- Alice J O’Toole, P Jonathon Phillips, Samuel Weimer, Dana A Roark, Julianne Ayyad, Robert Barwick, and Joseph Dunlop. Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision research*, 51(1):74–83, 2011.
- Liuba Papeo, Moritz F Wurm, Nikolaas N Oosterhof, and Alfonso Caramazza. The neural representation of human versus nonhuman bipeds and quadrupeds. *Scientific reports*, 7(1):14040, 2017.
- Marius V Peelen, Anthony P Atkinson, and Patrik Vuilleumier. Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, 30(30):10127–10134, 2010.
- Kevin A Pelphrey, James P Morris, Charles R Michelich, Truett Allison, and Gregory McCarthy. Functional anatomy of biological motion perception in posterior temporal cortex: an fmri study of eye, mouth and hand movements. *Cerebral cortex*, 15(12):1866–1876, 2005.
- Matthew A Lambon Ralph, Elizabeth Jefferies, Karalyn Patterson, and Timothy T Rogers. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1):42, 2017.
- Meike Ramon, Stephanie Caharel, and Bruno Rossion. The speed of recognition of personally familiar faces. *Perception*, 40(4):437–449, 2011.
- Constantin Rezlescu, Jason JS Barton, David Pitcher, and Bradley Duchaine. Normal acquisition of expertise with grebles in two cases of acquired prosopagnosia. *Proceedings of the National Academy of Sciences*, 111(14):5123–5128, 2014.
- Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27:169–192, 2004.
- Giacomo Rizzolatti and Corrado Sinigaglia. The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, 17(12):757, 2016.
- Giacomo Rizzolatti, Luciano Fadiga, Massimo Matelli, Valentino Bettinardi, Eraldo Paulesu, Daniela Perani, and Ferruccio Fazio. Localization of grasp representations in humans by pet: 1. observation versus execution. *Experimental brain research*, 111(2):246–252, 1996.
- Giacomo Rizzolatti, Leonardo Fogassi, and Vittorio Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews neuroscience*, 2(9):661, 2001.

- James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- James A Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805, 1999.
- James A Russell, Jo-Anne Bachorowski, and Jose-Miguel Fernandez-Dols. Facial and vocal expressions of emotion. *Annual review of psychology*, 54(1):329–349, 2003.
- Rebecca Saxe and Sean Dae Houlihan. Formalizing emotion concepts within a bayesian model of theory of mind. *Current opinion in Psychology*, 17:15–21, 2017.
- Rebecca Saxe and Nancy Kanwisher. People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Neuroimage*, 19(4):1835–1842, 2003.
- Annett Schirmer and Ralph Adolphs. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends in Cognitive Sciences*, 21(3): 216–228, 2017.
- Justine Sergent, Shinsuke Ohta, and BRENNAN MACDONALD. Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, 115(1):15–36, 1992.
- Thomas Serre, Aude Oliva, and Tomaso Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the national academy of sciences*, 104(15):6424–6429, 2007.
- Baoxu Shi and Tim Weninger. Open-world knowledge graph completion. 2018.
- Noa Simhi and Galit Yovel. The contribution of the body and motion to whole person recognition. *Vision research*, 122:12–20, 2016.
- Herbert A Simon. Information-processing theory of human problem solving. *Handbook of learning and cognitive processes*, 5:271–295, 1978.
- Amy E Skerry and Rebecca Saxe. A common neural code for perceived and inferred emotion. *Journal of Neuroscience*, 34(48):15997–16008, 2014.
- Amy E Skerry and Rebecca Saxe. Neural representations of emotion are organized around abstract event features. *Current biology*, 25(15):1945–1954, 2015.
- Julie S Snowden, JC Thompson, and D Neary. Knowledge of famous faces and names in semantic dementia. *Brain*, 127(4):860–872, 2004.
- Julie S Snowden, Jennifer C Thompson, and David Neary. Famous people knowledge and the right and left temporal lobes. *Behavioural neurology*, 25 (1):35–44, 2012.
- Robert Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. pages 4444–4451, 2017.

- Hugo J Spiers, Bradley C Love, Mike E Le Pelley, Charlotte E Gibb, and Robin A Murphy. Anterior temporal lobe tracks the formation of prejudice. *Journal of cognitive neuroscience*, 29(3):530–544, 2017.
- Ramprakash Srinivasan, Julie D Golomb, and Aleix M Martinez. A neural basis of facial action recognition in humans. *Journal of Neuroscience*, 36(16):4434–4442, 2016.
- Thomas K Srull and Robert S Wyer. Person memory and judgment. *Psychological review*, 96(1):58, 1989.
- Diana I Tamir and Mark A Thornton. Modeling the predictive social mind. *Trends in cognitive sciences*, 22(3):201–212, 2018.
- Joshua B Tenenbaum. Mapping a manifold of perceptual observations. pages 682–688, 1998.
- Mark A Thornton and Jason P Mitchell. Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral cortex*, 28(10):3505–3520, 2017.
- Mark A Thornton and Diana I Tamir. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, page 201616056, 2017.
- Mark A Thornton, Miriam E Weaverdyck, and Diana Tamir. The brain represents people as the mental states they habitually experience. *PsyArXiv*. August, 1, 2018.
- Simon Thorpe, Denis Fize, and Catherine Marlot. Speed of processing in the human visual system. *nature*, 381(6582):520, 1996.
- Alexander Todorov and James S Uleman. Spontaneous trait inferences are bound to actors’ faces: Evidence from a false recognition paradigm. *Journal of personality and social psychology*, 83(5):1051, 2002.
- Alexander Todorov and James S Uleman. The efficiency of binding spontaneous trait inferences to actors’ faces. *Journal of Experimental Social Psychology*, 39(6):549–562, 2003.
- Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5):675–691, 2005.
- Frank Tong, Ken Nakayama, Morris Moscovitch, Oren Weinrib, and Nancy Kanwisher. Response properties of the human fusiform face area. *Cognitive neuropsychology*, 17(1-3):257–280, 2000.
- Ernest C Tupes and Raymond E Christal. Recurrent personality factors based on trait ratings. *Journal of personality*, 60(2):225–251, 1992.
- Gilles Vannuscorps and Alfonso Caramazza. Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences*, 113(1):86–91, 2016.

- HL Wagner, CJ MacDonald, and AS Manstead. Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology*, 50(4):737, 1986.
- Yin Wang, Jessica A Collins, Jessica Koski, Tehila Nugiel, Athanasia Metoki, and Ingrid R Olson. Dynamic neural architecture for social knowledge retrieval. *Proceedings of the National Academy of Sciences*, 114(16):E3305–E3314, 2017.
- Christine E Watson, Eileen R Cardillo, Geena R Ianni, and Anjan Chatterjee. Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of cognitive neuroscience*, 25(8):1191–1205, 2013.
- Adam Waytz and Liane Young. Morality for us versus them. *Atlas of Moral Psychology*, page 186, 2018.
- Marlene Winell. Personal goals: The key to self-direction in adulthood. 1987.
- Yang Wu and Laura E Schulz. Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child development*, 89(2):649–662, 2018.
- Moritz Wurm and Alfonso Caramazza. Representation of action concepts in left posterior temporal cortex that generalize across vision and language. *bioRxiv*, page 361220, 2018.
- Moritz F Wurm and Angelika Lingnau. Decoding actions at different levels of abstraction. *Journal of Neuroscience*, 35(20):7727–7735, 2015.
- Moritz F Wurm, Giacomo Ariani, Mark W Greenlee, and Angelika Lingnau. Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cerebral Cortex*, 26(8):3390–3401, 2016.
- Moritz F Wurm, Alfonso Caramazza, and Angelika Lingnau. Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *Journal of Neuroscience*, 37(3):562–575, 2017.
- Eiling Yee, Michael N Jones, and Ken McRae. Semantic memory. *The Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience*, 3, 2018.
- Andrew W Young, Freda Newcombe, Edward HF de Haan, Marian Small, and Dennis C Hay. Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, 116(4):941–959, 1993.
- Galit Yovel and Alice J O’Toole. Recognizing people in motion. *Trends in Cognitive Sciences*, 20(5):383–395, 2016.
- Deborah Zaitchik, Caren Walker, Saul Miller, Pete LaViolette, Eric Feczko, and Bradford C Dickerson. Mental state attribution and the temporoparietal junction: an fmri study comparing belief, emotion, and perception. *Neuropsychologia*, 48(9):2528–2536, 2010.