

Opinion

Social Cognition 2.0: An Interactive Memory Systems Account

David M. Amodio^{1,2,*}

For 40 years, research on impression formation and attitudes has relied on dual-process theories that represent knowledge in a single associative network. Although such models explain priming effects and some implicit responses, they are generally silent on other forms of learning and on the interface of social cognition with perception and action. Meanwhile, advances in cognitive neuroscience reveal multiple, interacting forms of learning and memory (e.g., semantic associative memory, Pavlovian conditioning, and instrumental learning), with detailed models of their operations, neural bases, and connections with perceptual and behavioral systems. This memory systems perspective offers a more refined, neurally plausible model of social cognition and attitudes that, I argue, provides a useful and generative account of human social behavior.

Getting to Know You, Then and Now

How do you size up a new acquaintance? Is she friendly? Can you trust her? What are her likes and dislikes? And, as you get to know her – through direct interaction, observing her behavior, sharing emotional experiences, and hearing others' views – how does your impression evolve? This complex process is the subject of **social cognition** (see [Glossary](#)) – the study of how we form **representations** of other people and how these representations guide social perceptions, judgments, and actions.

For 40 years, social psychologists have relied on **dual-process models** to explain these social inferences. Dual-process models, which typically comprise an **automatic associative process** and a **more deliberative process**, have been useful for explaining conceptual priming, implicit **attitudes**, and judgment biases. Yet they are also limited; because they assume social information is represented within a single system of associative semantic memory, they do not discern among the many other ways we learn about people in everyday life. And when it comes to the role of **affect**, the influence of social cognition on perception, and the way social representations guide behavior, these models are often vague.

Meanwhile, since the initial emergence of dual-process models in social psychology, research on human learning and memory has marched on. Neurological case studies, in particular, provide powerful evidence for distinctions between forms of learning and memory – referred to as **memory systems** – such as **semantic** and **episodic** [1], **declarative** and **procedural** [2], and declarative and **Pavlovian** [3]. Contemporary cognitive neuroscience now reveals multiple, interacting systems of learning and memory, with increasingly detailed models of their distinct representations, operations, neural bases, and roles in judgment, perception, and action [4,5]. These advances, when applied to social cognition, suggest an opportunity to update our

Highlights

To date, research on impression formation and attitudes has relied on dual-process theories in which knowledge is represented in a single associative network.

Classic dual-process models explain priming effects and concept learning, but have trouble explaining how attitudes and impressions relate to affect, action, and perception.

Meanwhile, advances in cognitive neuroscience reveal multiple, interacting forms of learning and memory, with detailed models of their operations and neural bases.

This memory systems perspective offers a more refined, neurally plausible model of social cognition and attitudes that suggests a more precise and generative account of human social behavior.

¹Department of Psychology, New York University, New York, NY 10003, USA

²Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WT Amsterdam, The Netherlands

*Correspondence: david.amodio@nyu.edu (D.M. Amodio).

models of social cognition, from the single-system representational framework adopted in the 1970s to a contemporary model of interactive memory systems.

Social Cognition 1.0

Early research in social cognition sought to explain some quirks of human social perception: why do prior beliefs color one's impression and later recollection of a new acquaintance? **Why do we remember things better when they involve the self?** And why do egalitarians – people who consciously reject prejudice – sometimes show bias in their behavior? To explain these effects, social psychologists brought new ideas (at the time) from cognitive psychology such as categories, schemas, heuristics, priming, and automaticity [6–11] – ideas that, together, offered a new sociocognitive account of social inference [12].

Central to these advances was the information processing model (e.g., [13]), which holds that knowledge is stored in **a single associative (or connectionist) network in memory**, from which it may be activated automatically to influence thoughts and behaviors, independent of more deliberative judgment [13,14]. Placed in a dual-process framework, this **associative network** interacts with a second, more deliberative process, involving either **cognitive control** (Box 1), **propositional reasoning**, or **explicit beliefs**, which modulates or competes with the expression of the associative process [15–17] (Figure 1). Many so-called anomalies in social cognition – biased judgments, skewed impressions, and automatic stereotypes – could be explained by the interplay of these dual processes, whereby the second process fails to moderate the first (e.g., [15,18–23]).

For decades, this core engine of social cognition – the associative network – has been invoked to explain attitudes, trait impressions, stereotypes, and goals, among other phenomena [6,17–25]. Intervening years have seen important advances regarding the process through which this network operates, as well as the interplay of multiple representations within a single network [26,27]. However, when it comes to the fundamental issue of how information is represented, the assumption of a single associative system remains virtually unchanged.

Limitations of Single-System Representations

What's wrong with assuming a single system of associations? Although models that assume a **single system of knowledge representation explain some ways in which humans represent and use social knowledge**, when it comes to many other things humans do – experiencing and encoding affect, learning through direct interaction, and putting this knowledge into action – these models are often silent.

Box 1. Role of Cognitive Control

Although this article focuses on how attitudes and impressions are learned and represented, their expression is often modulated by cognitive control – a phenomenon that itself involves multiple processes [26,81,104,105]. For example, control may be initiated by the detection of conflict between an action and one's response goal, supported by the dorsal anterior cingulate cortex [106,107], or between one's action and external social cues, supported by the rostral paracingulate/medial prefrontal cortex [108]. Although control is typically assumed to be deliberative, the monitoring and detection of response conflict can operate without awareness [109].

Once the need for control is signaled, its implementation may involve different processes. For example, it may involve the abrogation of a response, linked to activity in the right inferior frontal gyrus [110], or the engagement of an intended response, supported by dorsal prefrontal cortex [111], potentially in conjunction with other processes that influence regulation (e.g., attention, emotion, and valuation). Moreover, the process of control may operate by overriding a bias – that is, by implementing an intended response without directly inhibiting the source of bias in the mind [112]. An important goal of future work is to understand how representations of impressions and attitudes, supported by their respective memory systems, interface with different components of control.

Glossary

Associative network: network of learned relationships between items (e.g., representing concepts, responses, or events) stored in long-term memory. Relationships vary in strength, and activation of one item spreads automatically to associated items.

Attitude: stable preference, representing a cognition ('I know X is good'), affective response (e.g., of pleasure), and/or behavioral disposition (e.g., to approach or avoid).

Declarative: refers to knowledge that can be explicitly accessed, subject to awareness, and directly reported. By contrast, nondeclarative associations operate outside of awareness and guide behavior implicitly; they are typically assessed with indirect measures.

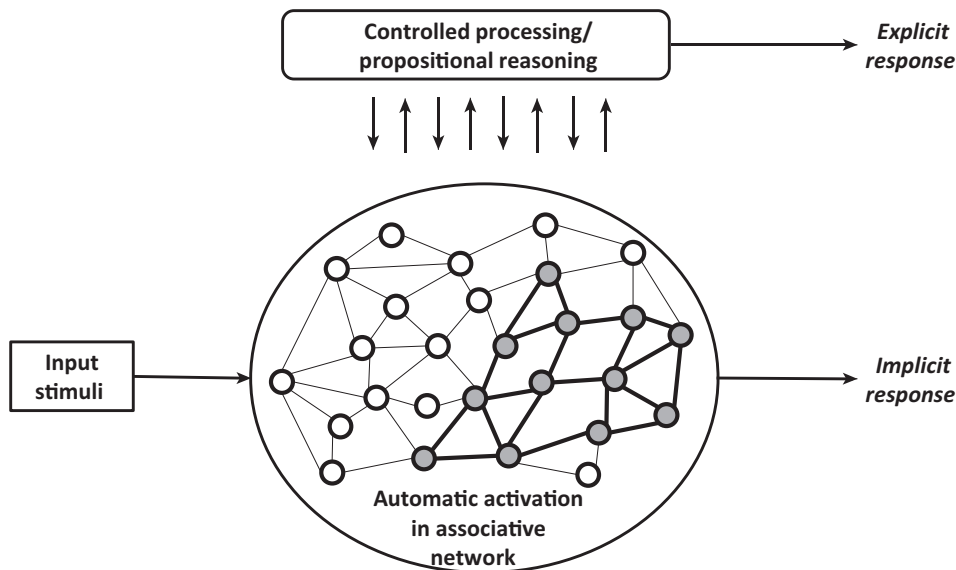
Dual-process model: theoretical model – common in social cognition to explain attitudes, social impressions, and decisions – that posits the operation of two processes. Most include (i) an associative process, in which knowledge stored in an associative network may be activated automatically and expressed implicitly, and (ii) a control process that modulates the expression of the associative process [e.g., 15,18–25,57]. Alternative models replace the control process with propositional validation [17] or a separate representation of explicit attitudes or beliefs [16].

Episodic memory: declarative knowledge of specific events in one's life, including the associated sensory and emotional experiences, timing, and context, which depends primarily on the hippocampus and surrounding medial temporal lobe.

Evaluation: current appraisal of preference, which influences choice.

Implicit associations: learned relationships among concepts, responses, and/or events that may be activated or expressed without awareness and probed with indirect assessments.

Instrumental learning (operant conditioning): goal-directed learning of actions that predict a reward, acquired through incremental reinforcement, expressed nondeclaratively, and supported



Trends in Cognitive Sciences

Figure 1. The Classic Dual-Process Model, with a Single Associative Representation. Most existing dual-process models of social cognition and attitudes include: (i) a single associative network (symbolic of connectionist) that represents stored knowledge, which may be automatically activated; and (ii) a regulatory process represented by either cognitive control or propositional reasoning. This type of model explains conceptual priming and automaticity effects, but does not account for different forms of learning and memory or their interface with physiological and behavioral responses. Adapted from [17].

Their limitations become conspicuous when considering neuropsychological studies of patients with selective brain lesions, where damage in a particular region impairs one learning capacity while sparing others [28]. Paired with neuroimaging and behavioral studies of learning and memory [29], this literature convincingly reveals the involvement of multiple interacting systems of learning and memory that likely support attitude and sociocognitive functions (Figure 2, Key Figure). Assuming there isn't a special brain module devoted to attitudes and impressions, separate from other kinds of memory, these findings strongly suggest the need to update our core model of social cognition.

Conceptual versus Affective Learning

Consider the case of affect: some models of implicit attitudes assume that evaluative associational networks are affective in nature (e.g., [17,26]). By this account, an attitude – the extent to which someone or something is liked or disliked – is represented by affective associations (i.e., feeling states), which then informs evaluative judgments and beliefs (i.e., degree of liking).

Yet, this theoretical position is complicated by classic studies of amygdala patients. In these studies, patients with selective amygdala damage and matched controls completed a simple Pavlovian conditioning task in which they viewed a series of stimuli [3,30,31]. During learning, conditioned stimuli (CSs) were paired with electrical shock. In a test phase, healthy control participants could report a conceptual **evaluation** ("the blue square is bad") while also exhibiting an aroused affective response to the CSs, indicated by a galvanic skin response. Amygdala lesion patients, by contrast, could form a conceptual evaluation but showed no evidence of affect [3,31], suggesting a role for the amygdala in affective learning corroborated by neuroimaging studies [32]. Moreover, a

primarily by ventral striatum and caudate.

Memory system: form of memory distinguished by its mode of learning, behavioral functions, and neural substrates. Major memory systems include semantic associative learning (i.e., priming), Pavlovian aversive or reward conditioning, instrumental learning, habit, semantic knowledge, and episodic memory.

Pavlovian conditioning (classical conditioning): learned association between a neutral stimulus and aversive outcome (in Pavlovian aversive conditioning, which depends on the amygdala) or rewarding outcome (in Pavlovian reward conditioning, which depends on the ventral pallidum and nucleus accumbens). The outcome stimulus is often described as biologically potent, in that it is intrinsically important to the organism and autonomically arousing.

Procedural memory: learned action associations acquired through incremental reinforcement, supported by the striatum and expressed nondeclaratively. Procedural memory comprises instrumental learning, skills (i.e., motor learning), and habits (persistent stimulus–response associations insensitive to change in reward contingency).

Representation: hypothetical array of nodes within an associative network that collectively produce a specific emergent mental construct (e.g., concepts and responses) or symbol of external reality (e.g., events).

Semantic associative memory (priming): learned relationships among concepts (e.g., regarding people, groups, and valence concepts) represented in a semantic network that may be activated automatically and expressed implicitly; primarily associated with anterior temporal lobe.

Semantic memory: declarative (i.e., explicit) knowledge of facts, concepts, and beliefs, primarily supported by lateral and ventral regions of the temporal lobe.

Social cognition: field of psychology concerned with the mental processes through which we perceive, think about, and act toward other people and in response to situational factors; also, a scientific

Key Figure

An Interactive Memory Systems Model

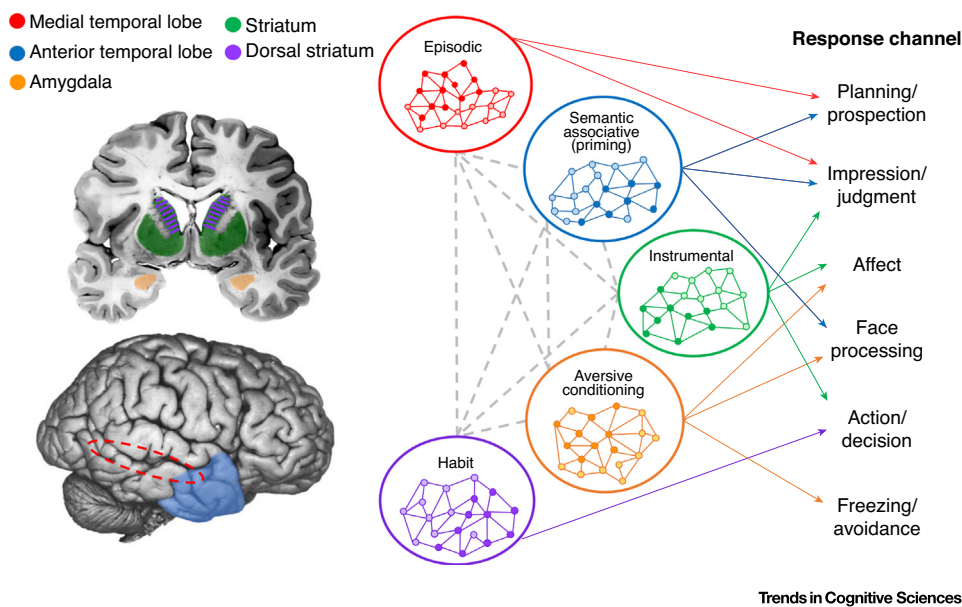


Figure 2. Cognitive neuroscience reveals multiple forms of learning and memory, associated with distinct neural substrates. A subset of these systems is illustrated here, including episodic memory, semantic associative memory (i. e., conceptual priming), instrumental (goal-directed) learning, Pavlovian aversive conditioning, habit, and their respective putative neural substrates (matched in color). These systems of memory may operate in concert, with distinct yet complementary operating parameters and functions, and are expressed through different subsets of response channels (e.g., among those shown here). This model suggests that we learn about people via multiple systems, encoding information in multiple representations, often simultaneously, and that these systems have complementary influences on judgments, decisions, and actions.

third group of participants, with selective hippocampal lesions, exhibited affective learning, indicated by galvanic skin response, in the absence of conceptual evaluation [33]. These results appear to reveal dissociable systems of conceptual and affective learning, rooted in different neural structures, which contribute to what social psychologists would call an attitude.

This dissociation – between conceptual and affective learning – concerns not just how we form attitudes, but also how we express them in behavior. In the Iowa gambling task, control subjects learn, through trial and error, to adjust their decision strategy following negative outcomes, and this change is foreshadowed by autonomic arousal to stimuli associated with large losses. However, when amygdala patients completed this task, they failed to show affective arousal and corresponding behavior change [33]. Without affective learning, their behavior was maladaptive. Considered together, these studies suggest separate conceptual and affective representations of an attitude, arising from separate declarative and Pavlovian memory systems, with different channels of expression. Although these systems operate in concert in the healthy brain, knowledge of these underlying demarcations improves our understanding of how an attitude is expressed in behavior.

Conceptual versus Instrumental Learning

Much of human social learning occurs via direct interaction – a colleague smiles when you say hello, suggesting she is friendly; another stares stoically ahead, suggesting he is not. Over time, you save your greetings for the friendly one. This pattern of reinforcement, in which a behavior is learned from feedback across many experiences, characterizes instrumental learning – a form of memory supported by the striatum that, along with skills and habits, constitutes procedural memory [34].

Studies of patients with Parkinson's disease (PD), associated with striatal dysfunction, reveal selective impairments in instrumental learning. For example, in probabilistic learning tasks, such as weather prediction [35] and probabilistic selection tasks [36], subjects must learn to choose stimuli that yield reward over others that do not. Over many learning trials, healthy subjects learn to choose the more frequently rewarded stimuli and report above-chance knowledge of reward contingencies. PD patients, by contrast, can report the correct reward contingencies, but remain at chance-level accuracy in their behavioral choices [35]. That is, they can form evaluations declaratively but not instrumentally. Moreover, hippocampal patients (i.e., amnesiacs) show the reverse impairment: their behavior reveals intact learning, yet they lack declarative knowledge of what they learned [35,37,38]. This double dissociation suggests independent representations of conceptual and instrumental attitude associations.

Beyond revealing different systems of attitude representation, these findings illuminate their complementary functions [39,40]. For example, different systems drive decisions under situations of high versus low cognitive load. With ample resources, the hippocampus and striatum jointly contribute to performance. But under load, hippocampal function is impaired while striatal function is spared, and thus the striatum primarily drives behavior [41]. This pattern suggests that declarative (i.e., conceptual) and instrumental representations operate in concert, adaptively, to guide performance amid different situational demands.

Timing matters, too: the striatum encodes feedback that immediately follows an action, but when feedback is delayed by even a few seconds, it becomes less responsive. By contrast, the hippocampus can encode delayed feedback, but is comparatively less involved when the feedback is immediate [42]. This pattern is supported by patient studies: PD patients can learn via delayed feedback but are impaired when feedback is immediate, whereas amnesiacs can learn from immediate feedback but are impaired when feedback is delayed [43]. Hence, depending on the timing of feedback, a different kind of representation – instrumental or declarative – will be formed and expressed. Again, though these systems normally function in concert, their dissociation reveals the true underlying structure of evaluative representations.

It is notable that a conventional dual-process model could be compatible with either the amygdala or PD patient findings described above, in that each distinguishes a declarative and nondeclarative process. But it can't explain both. These findings suggest (at minimum) a triple dissociation in evaluative learning – between hippocampal, amygdala, and striatal functions – a pattern demonstrated in rodent lesion studies [44] that has yet to be tested in humans.

Semantic Associative Memory

Of particular interest to sociocognitive theorists is the role of semantic associative memory (i.e., conceptual priming), which supports conceptual knowledge in an associative network that operates automatically and implicitly. Research on semantic associative memory suggests it is centered in the anterior temporal lobe (ATL) [45,46]. In functional magnetic resonance imaging (fMRI) studies, valence-based classification of words and faces has been associated with ATL

activity [47,48], often in conjunction with other regions involved in semantic activation (e.g., left inferior prefrontal cortex; [49]), and transcranial magnetic stimulation of the ATL, used to disrupt its activity, impairs semantic processing [50]. Although the ATL is not associated directly with affective processing, its connectivity with the amygdala and orbital frontal cortex suggests a link through which conceptual representations of valence interact with affective representations in other memory systems [51]. Semantic associative memory is further distinguishable from semantic knowledge, which is declarative, as well as from Pavlovian and instrumental learning.

Interacting Memory Systems

Although research often focuses on dissociations between memory systems, in an effort to delineate their respective functions and neural substrates, these systems typically operate in concert [4,52]. For example, in a social interaction, we simultaneously encode our affective response to a person, observe their appearance and behaviors, and learn from their responses to our own actions – a multimodal experience, recorded in systems responsive to different informational modalities. When forming a judgment, information from these memory systems converges, for example, in the dorsal medial frontal cortex, when inferring traits, or in the ventromedial prefrontal cortex when forming a choice, via connections to these regions from the amygdala, ATL, and striatum [53,54]. Hence, by identifying the interactive systems that represent impressions and attitudes and their influences on judgment and action, this approach is equipped to explain the complexities of impression and attitude formation and their roles in social judgments, perceptions, and behaviors (Box 2).

Implications for Theories of Attitudes and Social Cognition

The contemporary memory systems literature has important implications for social psychological theories of attitudes and social cognition. First and foremost, it suggests that models that assume single-system representations of associative knowledge – a feature of most existing dual-process models of attitudes and impression formation – are incomplete. There is now substantial and compelling evidence for multiple systems, with distinct operations and neural substrates that support different kinds of associative representations, experiences, and responses.

Second, most dual-process models in social psychology assume a single mode of associative learning, such that automatic associations are acquired through passive exposure to pairings of objects or events (stimulus–stimulus association). This focus on stimulus–stimulus associations is limited, however, in light of other known learning mechanisms, such as stimulus–outcome association (e.g., Pavlovian threat and reward) and action–outcome association (e.g., instrumental learning). Although stimulus–stimulus association explains one way in which attitudes and impressions are formed (e.g., in evaluative conditioning tasks) [55] and expressed (e.g., in

Box 2. Social Cognition Effects on Perception

A memory systems model also illuminates the influence of social cognition and attitudes on visual processing. For example, a threat response is known to heighten attention and visual acuity, via connections between the amygdala and regions of visual cortex [113–115]. This mechanism likely contributes to social threat responses [116], and when the threat is based in prejudice, it may influence the visual processing of faces biased racial perceptions [117,118]. Stereotypes may also influence visual perception by creating expectancies for configural processing. For example, the stereotype of Black hostility may facilitate the perception of anger on an African American man's face [119,120]. This effect is thought to represent semantic memory input to configural visual processing [53,121], supported by connectivity between the ATL and fusiform [122]. By considering memory system functions and their supporting neural circuitry, this approach offers a theoretical basis for how social cognition influences visual processing, further explaining how impressions and attitudes influence social perceptions and behavior.

priming tasks), a consideration of other learning mechanisms is needed to explain the roles of affect and behavior.

Third, an assumption of some dual-process models is that attitude associations are affective in nature (e.g., [17,56]). Yet, again, this view is disputed by memory systems research. Whereas conceptual associations (e.g., linking objects to valence concepts) are represented in semantic associative memory (i.e., priming), affective associations are represented by Pavlovian threat or reward associations. Because conventional implicit attitude tasks (e.g., priming tasks) assess associations between attitude objects and valence concepts without affording a direct assessment of affect (e.g., a measure of autonomic arousal; [57–59]), they likely reveal semantic associations, despite allusions to processes like Pavlovian or classical conditioning. Indeed, research designed to discern affective and conceptual associations in such tasks found evidence for conceptual priming but not affect [60]. To refer to such associations as affective may be a misnomer; they appear to be semantic. This distinction has critical implications for our understanding of how attitudes are learned and expressed, as described above.

Fourth, advocates of a single-process model posit that attitudes are formed via propositions [61,62]. Although the definition of proposition is broad, encompassing both causal and merely co-occurring relationships, a key tenet is that propositions hold truth value that is subjectively inferred and, hence, declarative [63,64]. However, substantial evidence from neuroscience – particularly neuropsychological studies, as described above – demonstrates nondeclarative learning. One compelling example comes from research on category learning – a form of procedural learning acquired through trial-and-error across many experiences. Amnesiacs often exhibit such learning, despite their inability to report it or to maintain working memory long enough to explicitly encode fluctuating probabilities of feedback across the task (e.g., [38,65,66]). And in work on aversive conditioning, associative learning has been shown in sea slugs (*Aplysia* [67]) and tiny roundworms (*C. elegans* [68]), organisms unlikely to engage in conscious propositional processing, through molecular mechanisms similar to those that support conditioning in humans [69]. From a modern learning and memory perspective, this single- versus dual-process debate seems moot; there are multiple forms of learning, some of which can operate without awareness. Moreover, although the memory systems model is incompatible with a single-process view, the opposite is not true: a memory systems model can account for propositional processing and add clarity to its role in evaluation and behavior.

Fifth, the memory systems model has implications for how implicit attitudes and impressions are updated, particularly in response to countervailing information. Early investigations suggested that **implicit associations** are resistant to change, whereas explicit impressions are updated rapidly [70,71]. However, recent work suggests that when new information is extremely countervailing (e.g., when Bob, who donates to charity and rescues pets, turns out to be a mafia hitman), implicit attitudes can change quickly [72,73]. Hence there is debate – do implicit associations change slowly or quickly? From a memory systems perspective, it depends on how an association is learned. Whereas aversive conditioning is resistant to change [74], semantic associations and instrumental reward associations can change, albeit slowly, with repeated experiences [25,75]. However, unlike semantic associations, which presumably weaken with repeated nonpairings between an object and its associated concept, instrumental reward updating is driven by prediction errors, and under some conditions, large prediction errors can cause rapid change [76]. Furthermore, some implicit attitude tasks may be more sensitive to the expression of one form of learning than another [77], further affecting interpretations of change. A memory systems account helps to explain when and how change is likely to occur while clarifying how such changes are expressed (and thus measured).

Finally, the memory systems model may help to illuminate the implicit–explicit distinction in social cognition and attitudes. First, this model identifies multiple forms of memory that support implicit (i.e., nondeclarative) learning [78,79]. In doing so, it clarifies that *implicit* is not a unitary phenomenon, but a broad descriptor that applies to a variety of learning mechanisms (e.g., Pavlovian conditioning, conceptual priming, and habit) which vary in function as well as in the way their operations may be considered implicit. This literature also distinguishes among explicit (i.e., declarative) memory processes (e.g., semantic and episodic memory). More broadly, a memory systems approach shifts focus from questions of implicitness to questions of function [4], and it aims to clarify not just whether a process is implicit, but how it is implicit and why this matters for impressions, attitudes, and behavior (see Outstanding Questions).

Toward Social Cognition 2.0: A Memory Systems Model of Attitudes and Social Cognition

The core model of social cognition, derived from 1970s cognitive psychology, is due for a fundamental system upgrade – a reboot based on contemporary knowledge of learning and memory as it relates to social impressions and attitudes. Some prior steps toward this upgrade are notable: Carlston's Associative Systems Theory [80] integrated ideas from social psychology, cognitive psychology, and neuroscience to propose multiple, interacting systems of verbal, affective, visual, and action representations that support social inference. And Lieberman *et al.* [81] proposed an elaboration of the classic dual-process model, differentiating between multiple processes that contribute to both automatic and controlled responses on the basis of emerging cognitive neuroscience findings – an approach used fruitfully to explain the phenomenon of intuition [82] and facets of the self concept [83]. In what follows, I describe how current thinking on learning and memory systems has begun to illuminate issues of implicit bias and impression formation.

Memory Systems Model of Implicit Bias

My colleagues and I initially adapted the memory systems approach to examine implicit prejudice and stereotyping – mental associations linking attitudes and traits, respectively, with social groups, which may influence behavior without awareness. Although well documented, the cognitive basis of implicit bias remains poorly understood, often defined vaguely as *associations in memory or traces of past experience* [84,85]. Such accounts assume a single system of implicit association and say little about how implicit biases are formed or operate. However, in light of the emerging literature on memory systems, we proposed that implicit biases could be represented by multiple forms of learning and memory, each with different implications for how implicit biases may be formed, expressed, and potentially reduced.

We first examined these implications as they pertained to affective and semantic forms of implicit racial bias, using a physiological index of amygdala activity to demonstrate an affective form of implicit bias that was independent of conceptual processing [57]. We then used this model to understand how different forms of racial associations – affective and semantic – relate to behavior. We found that separate measures of implicit evaluation (intended to assess affective associations) and stereotyping (assessing semantic associations) were uncorrelated and uniquely predicted different forms of discriminatory behavior [86]. To the extent this dissociation reflected separate semantic and Pavlovian systems, based on research linking implicit prejudice to the amygdala and Pavlovian conditioning [87,88], these findings suggested two different forms of implicit bias that should be expressed in different behaviors and changed via different kinds of interventions. A related distinction is made by the Dual Implicit Process Model [59], which notes that implicit attitudes could be represented by either negative

conceptual associations or threat associations – two forms of implicit attitude with different behavioral consequences, likely rooted in different memory systems.

The memory systems approach also offers predictions for how emotion, such as the anxiety people experience in intergroup situations, can influence the activation of implicit bias. Consistent with research relating fear and anxiety to the amygdala [89], but not to substrates of semantic association, we found that intergroup anxiety amplified subjects' implicit prejudiced attitudes but not their stereotype associations [90]. Of course, social interactions may involve any variety of emotions; the memory system offers a framework for understanding how they influence implicit attitudes and impressions.

Finally, the memory systems model posits that an implicit attitude could reflect associations in semantic memory (e.g., with the concept 'bad'), Pavlovian learning (e.g., affect), or instrumental memory (e.g., approach/avoidance tendencies). Whereas research linking implicit prejudice to amygdala activity is consistent with its basis in Pavlovian learning [57], an fMRI study using multivoxel pattern analysis revealed separate neural representations of implicit prejudice and stereotyping in the ATL, demonstrating how both could also exist within a single semantic network [48]. Further research is needed to understand the interplay of these implicit attitude components and their implications for behavior. Nevertheless, the memory systems approach has already begun to identify the specific neurocognitive processes that constitute implicit associations, moving the field beyond vague definitions, in an effort to elucidate their operation in the mind, behavior, and society.

Instrumental Learning of Attitudes and Trait Impressions

How do we form attitudes and impressions from direct social interaction? To date, research has examined how we learn about others vicariously or through observation, but not how we form impressions based on direct, dynamic social feedback. According to memory systems research, interaction-based impressions and attitudes depend on instrumental learning, rooted in action–outcome associations – that is, acting toward an interaction partner and learning from their response – in contrast to existing sociocognitive models assuming a single system of semantic associations.

Instrumental learning, often examined using probabilistic reward reinforcement paradigms and supported by the striatum, has been shown to guide choice preferences [92,93]. Although not typically interpreted in terms of an attitude, such effects resemble early conceptualizations of attitudes as response dispositions – the behavioral component of classic tripartite models [58,94] (Box 3). Moreover, because instrumental reward learning involves action associations [95,96], an instrumental-based attitude may have more direct effects on behavior than attitudes represented by propositions or semantic associations [76,97]. An instrumental learning account of attitudes also suggests a mechanism for the formation of habits – actions that persist after reward has ceased [98,99].

The role of instrumental learning in impression formation was examined in a recent experiment [100]. In this study, participants played a money-sharing game; on each trial, they chose to interact with one of two players and then received feedback on the amount that player shared. Importantly, players shared a large or small amount of money from either a large or small endowment, such that they were associated with either high or low reward independent of their degree of generosity. Computational modeling of prediction errors revealed that representations of instrumental reward value and trait generosity were encoded separately. Although both kinds of prediction errors correlated with ventral striatum activity, generosity prediction errors

Box 3. Revisiting the Tripartite Model of Attitudes

The model of attitudes suggested here – involving separate, interacting representations of conceptual, affective, behavioral associations – may sound familiar. Indeed, the classic tripartite theory in social psychology posits that an attitude reflects some combination of beliefs, affective responses, and behavioral tendencies regarding the attitude object [58,123]. However, empirical support for these distinctions was hampered by methodological limitations, and the attitude concept has become increasingly associated with affect, relative to cognition, and rarely with behavior [58].

The memory systems model described here parallels this classic view, but with some critical updates and differences. First, it suggests a plausible basis for the behavioral component of attitudes in instrumental learning. Although instrumental learning appeared in earlier proposals [58,91], the current literature offers a significantly advanced model of instrumental learning, with more precise operationalizations and new methods to study it. Second, it expands on the tripartite components, such that they likely represent multiple underlying systems. For example, the affective component could reflect Pavlovian reward or aversive conditioning; the cognitive component could reflect episodic or semantic knowledge or semantic associative memory; and the behavioral component could reflect goal-directed instrumental or skill learning.

That said, it is difficult to classify memory systems as strictly relating to cognitive, affective, or behavioral components of an attitude, given the more specific and interactive functions of these systems. For example, although instrumental learning relates most directly to action, other systems also play a role: Pavlovian fear conditioning produces freezing, whereas Pavlovian reward conditioning can prepare an animal for approach behavior, and semantic associations, along with semantic and episodic knowledge, support the planning and guidance of a response. Hence, a memory systems model complements and supports the classic tripartite model, illuminating underlying processes that give rise to major attitude components and providing an expanded theoretical model of attitude formation, expression, conflict, and change.

were also correlated with regions previously linked to explicit trait updating (e.g., right temporoparietal junction, precuneus, and interparietal lobule; [101]). Furthermore, reward and generosity representations had unique effects on subsequent social choices.

These findings reveal that instrumental learning can support the formation of both attitudes and trait impressions from direct social interaction. Moreover, trait learning in social interactions may depend on the interplay of instrumental and conceptual systems – that is, social information encoded simultaneously through reinforcement and observation. Together, this work demonstrates a basis for implicit attitudes in instrumental learning, apparently independent of conceptual association or Pavlovian conditioning, and consistent with prior research on probabilistic reinforcement learning that functions independently of declarative knowledge [38,41].

More broadly, the emerging body of memory systems research is revealing a more comprehensive account of how we interact with people in everyday life – from our first impressions to our evolving attitudes, formed vicariously or in person, and expressed in our perceptions, judgments, emotions, and behavior (Figure 2). By focusing on psychological function, rather than implicitness/explicitness, and with a consideration of neural mechanism and the situational affordances that affect learning and behavior, the memory systems approach is beginning to illuminate the ‘black box’ portrayed by the single-system representational account of current dual-process models.

Concluding Remarks

Our knowledge of learning and memory has advanced considerably since dual-process models of attitudes and social cognition first emerged. These advances offer mounting evidence for an interactive memory systems account of social cognition and attitudes, particularly as they relate to the implicit associations that drive impressions, preferences, perceptions, and behaviors. This memory systems account is more consistent with neural function than are dual-process

Outstanding Questions

How exactly do different memory systems interact with each other? Do they interact directly, or in higher-level representations (supported by associational structures in the brain)? Do memory systems always operate in concert or do they sometimes compete during learning or response formation?

How does change in one type of representation relate to change in another? Are changes always concurrent and congruent?

Is it possible to have inconsistent representations, which may suggest new explanations for attitude conflict and mixed trait impressions? How do conflicting representations relate to behavior?

Are there individual differences in people's reliance on a particular memory system for the formation and expression of attitudes and impressions?

What are the implications of a memory systems model for interventions, for example, to reduce implicit racial bias? Are different interventions required to alter different representations (e.g., in semantic vs instrumental associations)?

accounts, addressing a wider range of responses, and it provides substance to the vague notion of *associations in memory* invoked by dominant theories [79,84]. In light of recent advances in social psychology and cognitive neuroscience, along with growing concern about dual-process frameworks [59,99,102–104], the time is ripe for an upgrade: from traditional dual process to a model of interacting memory systems – an approach that offers a rich theoretical foundation to understand the complexities of social cognition and social behavior.

Acknowledgments

This article benefited from the generous and insightful feedback of Don Carlston, Karin Foerde, and members of the NYU Social Neuroscience Laboratory (Jeffrey Berg, Michael Berkebile, Hope Oloye, and Ben Stillerman) and support from the US National Science Foundation (BCS 1551826) and the Netherlands Organisation for Scientific Research (VICI 016.185.058).

References

1. Tulving, E. (1985) How many memory systems are there? *Am. Psychol.* 40, 385–398
2. Squire, L.R. and Zola-Morgan, S. (1988) Memory: brain systems and behavior. *Trends Neurosci.* 11, 170–175
3. Bechara, A. *et al.* (1995) Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* 269, 1115–1118
4. Henke, K. (2010) A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* 11, 523–532
5. Squire, L.R. (2004) Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177
6. Bargh, J.A. (1982) Attention and automaticity in the processing of self-relevant information. *J. Pers. Soc. Psychol.* 43, 425–436
7. Hastie, R. and Kumar, P.A. (1979) Person memory: personality traits as organizing principles in memory for behaviors. *J. Pers. Soc. Psychol.* 37, 25–38
8. Higgins, E.T. *et al.* (1977) Category accessibility and impression formation. *J. Exp. Soc. Psychol.* 13, 141–154
9. Kahneman, D. and Tversky, A. (1973) On the psychology of prediction. *Psychol. Rev.* 80, 237–251
10. Langer, E.J. *et al.* (1978) The mindlessness of ostensibly thoughtful action: the role of “placebic” information in interpersonal interaction. *J. Pers. Soc. Psychol.* 36, 635–642
11. Wegner, D.M. and Vallacher, R.R. (1977) *Implicit Psychology: An Introduction to Social Cognition*, Oxford University Press
12. Wyer, R.S. and Carlston, D.E. (1979) *Social Cognition, Inference, and Attribution*, Psychology Press
13. Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84, 1–66
14. Posner, M. and Snyder, C. (1975) Attention and cognitive control. In *Information Processing and Cognition: The Loyola Symposium* (Solo, R.L., ed.), pp. 55–85, Erlbaum
15. Devine, P.G. (1989) Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 518
16. Wilson, T.D. *et al.* (2000) A model of dual attitudes. *Psychol. Rev.* 107, 101–126
17. Gawronski, B. and Bodenhausen, G.V. (2006) Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychol. Bull.* 132, 692–731
18. Gilbert, D.T. *et al.* (1988) On cognitive busyness: when person perceivers meet persons perceived. *J. Pers. Soc. Psychol.* 54, 733–740
19. Gilbert, D.T. and Hixon, J.G. (1991) The trouble of thinking: activation and application of stereotypic beliefs. *J. Pers. Soc. Psychol.* 60, 509–517
20. Brewer, M.B. (1988) *A Dual Process Model of Impression Formation*, Lawrence Erlbaum Associates
21. Fiske, S.T. and Pavelchak, M.A. (1986) Category-based versus piecemeal-based affective responses: developments in schema-triggered affect. In *Handbook of Motivation and Cognition: Foundations of Social Behavior* (Sorrentino, R.M. and Higgins, E.T., eds), pp. 167–203, Guilford Press
22. Trope, Y. (1986) Identification and inferential processes in dispositional attribution. *Psychol. Rev.* 93, 239–257
23. Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131
24. Smith, E.R. and DeCoster, J. (1998) Knowledge acquisition, accessibility, and use in person perception and stereotyping: simulation with a recurrent connectionist network. *J. Pers. Soc. Psychol.* 74, 21–35
25. Smith, E.R. and DeCoster, J. (2000) Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Pers. Soc. Psychol. Rev.* 4, 108–131
26. Cunningham, W.A. and Zelazo, P.D. (2007) Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends Cogn. Sci.* 11, 97–104
27. Sherman, J.W. *et al.* (2008) The self-regulation of automatic associations and behavioral impulses. *Psychol. Rev.* 115, 314–335
28. Gabrieli, J.D. (1998) Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* 49, 87–115
29. Poldrack, R.A. and Packard, M.G. (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41, 245–251
30. LaBar, K.S. *et al.* (1995) Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.* 15, 6846–6855
31. Phelps, E.A. *et al.* (1998) Specifying the contributions of the human amygdala to emotional memory: a case study. *Neurocase* 4, 527–540
32. LaBar, K.S. *et al.* (1998) Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20, 937–945
33. Bechara, A. *et al.* (1999) Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481
34. Yin, H.H. and Knowlton, B.J. (2006) The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476
35. Knowlton, B.J. *et al.* (1996) A neostriatal habit learning system in humans. *Science* 273, 1399–1402
36. Frank, M.J. *et al.* (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306, 1940–1943
37. Reber, P.J. and Squire, L.R. (1994) Parallel brain systems for learning with and without awareness. *Learn. Mem.* 1, 217–229
38. Knowlton, B.J. *et al.* (1994) Probabilistic classification learning in amnesia. *Learn. Mem.* 1, 106–120

39. Foerde, K. and Shohamy, D. (2011) The role of the basal ganglia in learning and memory: insight from Parkinson's disease. *Neurobiol. Learn. Mem.* 96, 624–636
40. Poldrack, R.A. *et al.* (2001) Interactive memory systems in the human brain. *Nature* 414, 546–550
41. Foerde, K. *et al.* (2006) Modulation of competing memory systems by distraction. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11778–11783
42. Foerde, K. and Shohamy, D. (2011) Feedback timing modulates brain systems for learning in humans. *J. Neurosci.* 31, 13157–13167
43. Foerde, K. *et al.* (2013) A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. *J. Neurosci.* 33, 5698–5704
44. McDonald, R.J. and White, N.M. (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behav. Neurosci.* 107, 3–22
45. Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45
46. Ralph, M.A.L. *et al.* (2017) The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55
47. Cunningham, W.A. *et al.* (2004) Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J. Cogn. Neurosci.* 16, 1717–1729
48. Gilbert, S.J. *et al.* (2012) Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia* 50, 3600–3611
49. Poldrack, R.A. *et al.* (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35
50. Pobric, G. *et al.* (2007) Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20137–20141
51. Olson, I.R. *et al.* (2013) Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc. Cogn. Affect. Neurosci.* 8, 123–133
52. Cabeza, R. and Moscovitch, M. (2013) Memory systems, processing modes, and components: functional neuroimaging evidence. *Perspect. Psychol. Sci.* 8, 49–55
53. Amodio, D.M. (2014) The neuroscience of prejudice and stereotyping. *Nat. Rev. Neurosci.* 15, 670–682
54. de Schotten, M.T. *et al.* (2012) Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex* 48, 82–96
55. Olson, M.A. and Fazio, R.H. (2001) Implicit attitude formation through classical conditioning. *Psychol. Sci.* 12, 413–417
56. Fazio, R.H. (1990) Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. In *Advances in Experimental Social Psychology*, pp. 75–109, Elsevier
57. Amodio, D.M. *et al.* (2003) Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *J. Pers. Soc. Psychol.* 84, 738–753
58. Breckler, S.J. (1984) Empirical validation of affect, behavior, and cognition as distinct components of attitude. *J. Pers. Soc. Psychol.* 47, 1191–1205
59. March, D.S. *et al.* (2018) Clarifying the explanatory scope of the dual implicit process model. *Psychol. Inq.* 29, 38–43
60. Blaison, C. *et al.* (2012) The affect misattribution procedure: hot or not? *Emotion* 12, 403–412
61. De Houwer, J. (2009) The propositional approach to associative learning as an alternative for association formation models. *Learn. Behav.* 37, 1–20
62. Kruglanski, A.W. and Thompson, E.P. (1999) Persuasion by a single route: a view from the unimodel. *Psychol. Inq.* 10, 83–109
63. Dawson, M.E. *et al.* (2007) Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion* 7, 755–766
64. De Houwer, J. (2018) Propositional models of evaluative conditioning. *Soc. Psychol. Bull.* 13, e28046
65. Reber, P.J. *et al.* (1996) Dissociable properties of memory systems: differences in the flexibility of declarative and non-declarative knowledge. *Behav. Neurosci.* 110, 861–871
66. Nagy, H. *et al.* (2007) Cognitive sequence learning in Parkinson's disease and amnesic mild cognitive impairment: dissociation between sequential and non-sequential learning of associations. *Neuropsychologia* 45, 1386–1392
67. Carew, T. *et al.* (1983) Differential classical conditioning of a defensive withdrawal reflex in *Aplysia californica*. *Science* 219, 397–400
68. Ardiel, E.L. and Rankin, C.H. (2010) An elegant mind: learning and memory in *Caenorhabditis elegans*. *Learn. Mem.* 17, 191–201
69. Lynch, M.A. (2004) Long-term potentiation and memory. *Physiol. Rev.* 84, 87–136
70. Gregg, A.P. *et al.* (2006) Easier done than undone: asymmetry in the malleability of implicit preferences. *J. Pers. Soc. Psychol.* 90, 1–20
71. Rydell, R.J. and McConnell, A.R. (2006) Understanding implicit and explicit attitude change: a systems of reasoning analysis. *J. Pers. Soc. Psychol.* 91, 995–1008
72. Cone, J. and Ferguson, M.J. (2015) He did what? The role of diagnosticity in revising implicit evaluations. *J. Pers. Soc. Psychol.* 108, 37–57
73. Mann, T.C. and Ferguson, M.J. (2015) Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *J. Pers. Soc. Psychol.* 108, 823–849
74. Bouton, M.E. (2004) Context and behavioral processes in extinction. *Learn. Mem.* 11, 485–494
75. Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning: An Introduction*, MIT Press
76. Hackel, L.M. and Amodio, D.M. (2018) Computational neuroscience approaches to social cognition. *Curr. Opin. Psychol.* 24, 92–97
77. Gawronski, B. and Hahn, A. (2019) Implicit measures: procedures, use and interpretation. In *Measurement in Social Psychology* (Blanton, J.M.L.H. and Webster, G.D., eds), Taylor & Francis
78. Squire, L.R. and Zola, S.M. (1996) Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13515–13522
79. Amodio, D.M. and Ratner, K.G. (2011) A memory systems model of implicit social cognition. *Curr. Dir. Psychol. Sci.* 20, 143–148
80. Carlston, D.E. (1994) *Associated Systems Theory: A Systematic Approach to Cognitive Representations of Persons*, Lawrence Erlbaum Associates
81. Lieberman, M.D. *et al.* (2002) Reflexion and reflection: a social cognitive neuroscience approach to attributional inference. In *Advances in Experimental Social Psychology*, pp. 199–249, Elsevier
82. Lieberman, M.D. (2000) Intuition: a social cognitive neuroscience approach. *Psychol. Bull.* 126, 109–137
83. Lieberman, M.D. *et al.* (2004) Evidence-based and intuition-based self-knowledge: an fMRI study. *J. Pers. Soc. Psychol.* 87, 421–435
84. Greenwald, A.G. and Banaji, M.R. (2017) The implicit revolution: reconceiving the relation between conscious and unconscious. *Am. Psychol.* 72, 861–871
85. Greenwald, A.G. and Banaji, M.R. (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27
86. Amodio, D.M. and Devine, P.G. (2006) Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *J. Pers. Soc. Psychol.* 91, 652–661

87. Cunningham, W.A. *et al.* (2004) Separable neural components in the processing of black and white faces. *Psychol. Sci.* 15, 806–813
88. Olsson, A. *et al.* (2005) The role of social groups in the persistence of learned fear. *Science* 309, 785–787
89. Davis, M. (1992) The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* 15, 353–375
90. Amodio, D.M. and Hamilton, H.K. (2012) Intergroup anxiety effects on implicit racial evaluation and stereotyping. *Emotion* 12, 1273–1280
91. Insko, C.A. (1965) Verbal reinforcement of attitude. *J. Pers. Soc. Psychol.* 2, 621–623
92. O'Doherty, J.P. *et al.* (2017) Learning, reward, and decision making. *Annu. Rev. Psychol.* 68, 73–100
93. Ruff, C.C. and Fehr, E. (2014) The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* 15, 549–562
94. Eagly, A.H. and Chaiken, S. (1998) Attitude structure and function. In *The Handbook of Social Psychology*. Oxford University Press
95. Rescorla, R.A. and Solomon, R.L. (1967) Two-process learning theory: relationships between Pavlovian conditioning and instrumental learning. *Psychol. Rev.* 74, 151–182
96. Samejima, K. *et al.* (2005) Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340
97. Amodio, D.M. and Berg, J.J. (2018) Toward a multiple memory systems model of attitudes and social cognition. *Psychol. Inq.* 29, 14–19
98. Foerde, K. (2018) What are habits and do they depend on the striatum? A view from the study of neuropsychological populations. *Curr. Opin. Behav. Sci.* 20, 17–24
99. Wood, W. (2017) Habit in personality and social psychology. *Pers. Soc. Psychol. Rev.* 21, 389–403
100. Hackel, L.M. *et al.* (2015) Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nat. Neurosci.* 18, 1233–1235
101. Mende-Siedlecki, P. *et al.* (2013) Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *J. Neurosci.* 33, 19406–19415
102. Ferguson, M.J. *et al.* (2014) Rethinking duality: criticisms and ways forward. In *Dual-Process Theories of the Social Mind* (Sherman, J.W. *et al.*, eds), pp. 578–594, Guilford Press
103. Melnikoff, D.E. and Bargh, J.A. (2018) The mythical number two. *Trends Cogn. Sci.* 22, 280–293
104. Amodio, D.M. (2014) Dual experiences, multiple processes: looking beyond dualities for mechanisms of the mind. In *Dual-Process Theories of the Social Mind* (Sherman, J.W. *et al.*, eds), pp. 560–577, Guilford Press
105. Amodio, D.M. *et al.* (2004) Neural signals for the detection of unintentional race bias. *Psychol. Sci.* 15, 88–93
106. Kerns, J.G. *et al.* (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026
107. Shenhav, A. *et al.* (2016) Dorsal anterior cingulate cortex and the value of control. *Nat. Neurosci.* 19, 1286–1291
108. Amodio, D.M. and Frith, C.D. (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277
109. Berns, G.S. *et al.* (1997) Brain regions responsive to novelty in the absence of awareness. *Science* 276, 1272–1275
110. Aron, A.R. *et al.* (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177
111. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
112. Amodio, D.M. and Swencionis, J.K. (2018) Proactive control of implicit bias: A theoretical model and implications for behavior change. *J. Pers. Soc. Psychol.* 115, 255–275
113. Armony, J.L. and Dolan, R.J. (2002) Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. *Neuropsychologia* 40, 817–826
114. Öhman, A. *et al.* (2001) Emotion drives attention: detecting the snake in the grass. *J. Exp. Psychol. Gen.* 130, 466–478
115. Phelps, E.A. *et al.* (2006) Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol. Sci.* 17, 292–299
116. Öhman, A. (1986) Face the beast and fear the face: animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology* 23, 123–145
117. Dotsch, R. *et al.* (2008) Ethnic out-group faces are biased in the prejudiced mind. *Psychol. Sci.* 19, 978–980
118. Ofan, R.H. *et al.* (2011) Seeing race: N170 responses to race and their relation to automatic racial attitudes and controlled processing. *J. Cogn. Neurosci.* 23, 3153–3161
119. Hugenberg, K. and Bodenhausen, G.V. (2003) Facing prejudice: implicit prejudice and the perception of facial threat. *Psychol. Sci.* 14, 640–643
120. Kawakami, K. *et al.* (2017) Intergroup perception and cognition: an integrative framework for understanding the causes and consequences of social categorization. In *Advances in Experimental Social Psychology*, pp. 1–80, Elsevier
121. Stoller, R.M. and Freeman, J.B. (2016) Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* 19, 795–797
122. Nestor, A. *et al.* (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9998–10003
123. Rosenberg, M.J. and Hovland, C.I. (1966) Attitude organization and change: an analysis of consistency among attitude components. In *Attitude Organization and Change: An Analysis of Consistency among Attitude Components* (Rosenberg, M.J. and Hovland, C.I., eds), Yale University Press