nature reviews psychology

https://doi.org/10.1038/s44159-025-00445-x

Perspective

Check for updates

A learning and memory account of impression formation and updating

David M. Amodio

Abstract

Impression formation is the process of learning about people — how a perceiver infers another person's traits, goals and preferences while also forming their own attitude towards that person. Emerging research shows that impression formation involves a variety of learning mechanisms — a multimodal process rooted in multiple underlying memory systems. In this Perspective, I describe the roles of episodic, semantic, instrumental and Pavlovian memory systems in impression formation and updating. By considering the unique and interactive functions of learning and memory mechanisms, this memory systems framework expands and clarifies theories of how impressions are formed, changed and expressed in behaviour, moving beyond prior accounts based on semantic memory models. This framework also illuminates longstanding debates on the nature of implicit social cognition and how social information is represented in the mind.

Sections

Department of Psychology, University of Amsterdam, Amsterdam, Netherlands. e-mail: d.m.amodio@uva.nl

Introduction

Humans depend on other people to survive and thrive, and the ability to assess others — to infer their traits and motives; to discern friend from foe — is a fundamental capacity of the human mind[1]. This capacity is known as impression formation, and while a central topic of social cognition research, it reflects the culmination of many basic cognitive, perceptual and affective processes studied across areas of psychology.

Social impressions have long been considered multifaceted[2–4] — they involve conceptual knowledge of a target person's attributes, such as their trait characteristics[1,5–7], goals and intentions[8–11], mental states[12,13] and life circumstances (for example, wealth, geography or group memberships)[14–16]. Impressions also involve a perceiver's attitude towards a target[17–19], which can include the perceiver's evaluative beliefs and associations, affective responses and behavioural dispositions (for example, to approach or avoid)[20]. These varied aspects of an impression reflect the multiple ways in which humans learn and how these learning processes function together to guide social judgements, decisions and actions[21]. These learning mechanisms further guide how impressions are changed (updated) in response to new information and experiences.

The notion that human thought and behaviour are rooted in mechanisms of learning and memory is foundational in psychological science[22–28], and it is this idea that inspired the emergence of social cognition — a field originally known as 'person memory'[29]. Human learning and memory can be understood as a set of interacting memory systems, each characterized by a unique profile of operation, psychological function, mode of expression and neural substrate[30–33], that operate in concert to support adaptive behaviour[21,34,35]. Although classic accounts of impression formation emphasize conceptual inferences rooted in semantic memory[1,17,29,36], it is now clear that multiple learning and memory mechanisms contribute to how people think about and act towards others.

In this Perspective, I describe major mechanisms of learning and memory that support impression formation, integrating theory and research from social psychology, cognitive psychology and neuroscience. I begin by describing key learning and memory processes involved in social cognition — episodic memory, semantic memory, instrumental learning and Pavlovian learning — and discuss their implications for impression formation and updating. I then discuss how a consideration of these learning and memory systems and their interactions illuminate longstanding theoretical questions regarding the nature of implicit attitudes and process models of social cognition.

Learning and memory systems

When an individual meets someone, they experience that person in multiple ways simultaneously: they encode the details of the event (such as the person's appearance and nonverbal attributes, the other people involved and the context), they infer the person's trait attributes and goals, they track whether the person responds positively or negatively to the things they do and say, and they react emotionally to the person (Fig. 1). Each form of learning — the details encoded (episodic memory), the traits inferred (semantic memory), the feedback tracked (instrumental learning) and the affect experienced (Pavlovian learning) — contributes to the emergent impression formed of that person.

In this section, I describe the major learning and memory systems that are most relevant to impression formation. I highlight the specific kind of information each system encodes, how this information is typically expressed, the degree to which it is consciously accessible and how it is updated. I also note the neural substrates of different learning mechanisms to illustrate their functional separation and connections to cognitive processes underlying judgement and behaviour (see Box 1 for a similar memory systems analysis of intergroup social cognition).

Semantic memory

Semantic memory refers to the learning, representation and retrieval of general knowledge — 'the sky is blue', '1 + 1 = 2' or 'Sally is friendly, clever and athletic'. Early theories of person perception and social cognition were inspired by models of semantic memory[29,37], and

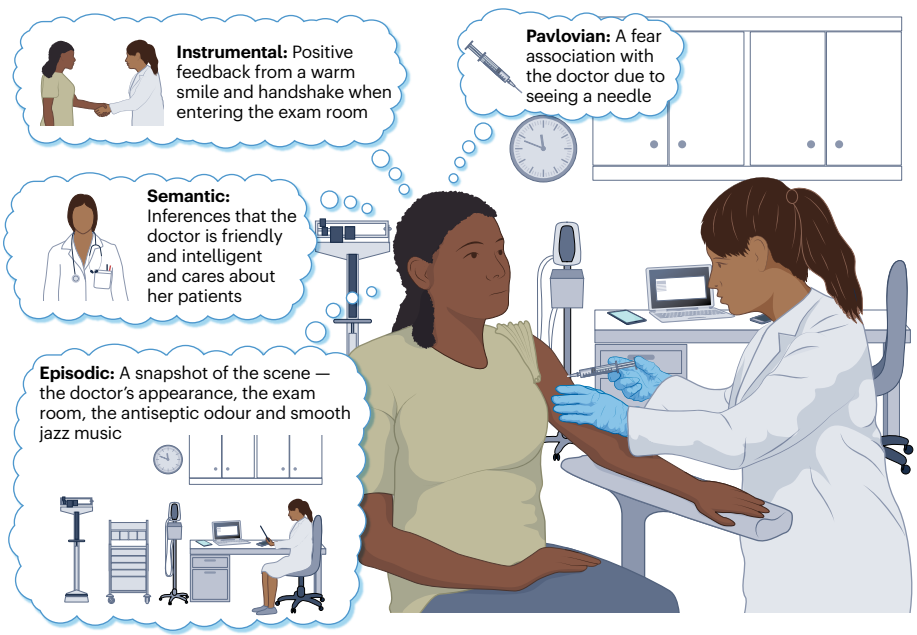

Fig. 1 | Contributions of different learning and memory systems to impression formation. As a perceiver forms an impression, they simultaneously encode information through multiple memory systems. For example, when meeting a doctor for a vaccination, a person might infer their traits as intelligent and caring (semantic), form behavioural approach associations from their positive feedback (instrumental), and form a physiological fear association with them upon seeing the needle nearby (Pavlovian), all while encoding the multimodal details of the situation (episodic).

Box 1 | Memory systems and intergroup social cognition

Impression formation and intergroup social cognition are closely related phenomena: whereas impression formation concerns individual-level processes, intergroup research extends this scope to include group and societal levels of analysis, with group-level traits and preferences corresponding to stereotypes and prejudice. Thus, a memory systems analysis of impression formation also informs theories of stereotyping and prejudice[211].

Clarifying representations of intergroup bias

From a memory systems perspective, and in line with classic theories[54], stereotypes reflect knowledge in semantic memory which can be expressed directly in verbal reports or indirectly in conceptual word classifications. However, a memory systems analysis departs from classic theories by identifying multiple forms of prejudice (that is, a group-level attitude). These correspond to semantic evaluative associations, instrumental reward associations and Pavlovian threat associations — each of which reflects a form of group-based preference. Because an individual's intergroup bias could involve any combination of these processes, a memory systems framework accounts for why prejudiced attitudes do not always align with stereotype knowledge and why some forms of intergroup bias are more likely than others to produce discriminatory behaviour[142,156,169].

Measuring intergroup bias

A memory systems framework also illuminates the measurement of intergroup bias. For example, it clarifies that word-based implicit-prejudice tasks primarily assess semantic evaluation and that other measures are needed to assess affective or motivational (that is, instrumental) processes. It further suggests that self-report and implicit measures might differ not only because of the different task design features[212], but also because they afford expressions of different underlying processes[111,125,139,213].

Predicting bias in behaviour

A common critique of intergroup research is that implicit bias measures are often weakly associated with behaviour. The memory systems approach was developed, in part, to clarify how measures of implicit bias should predict behaviour[168]. It suggests that measures of bias reflecting semantic associations should primarily predict high-level judgements and verbal behaviour, whereas measures reflecting instrumental or Pavlovian associations should be more predictive of nonverbal behaviours (for example, approach or social distancing).

Reducing prejudice

A memory systems analysis informs prejudice reduction by clarifying which aspects of bias are changeable and how they might be changed. By considering mechanisms of updating, it informs the design of interventions and assessment of their impacts. Furthermore, by identifying forms of prejudice that are difficult to change (such as Pavlovian associations or habits), this analysis highlights the importance of structural interventions that supersede individual-level responses[214]. That is, a memory systems framework suggests that although models of individual-level processes are critical for understanding how biases are formed and expressed, the effective reduction of bias often requires structural-level interventions.

most contemporary models of impression formation continue to assume a basis in semantic processes[21,38].

Semantic memory is declarative, such that it is explicitly reportable, and propositional, in that it meaningfully links abstract linguistic concepts[39]. Semantic memory is primarily represented in the anterior temporal lobe[40,41] and activated during social judgements in the medial frontal cortex[42,43]. Although typically expressed via verbal self-report, semantic associations can also be expressed on indirect measures that involve conceptual categorizations (for example, semantic priming). That is, although a perceiver is aware of their semantic knowledge, this knowledge can be expressed indirectly (that is, implicitly) and therefore potentially without one's intention or awareness.

In the context of impression formation, semantic memory supports knowledge regarding a person's traits, goals, circumstances and evaluation[1,10,11,17], encoded as cognitive concepts organized in a semantic network[22,44]. Semantic impressions can be based on direct verbal descriptions of a person[1] or inferred from a person's behaviour[3,7,45,46].

When semantic knowledge is activated, such as when a target individual is encountered, this information becomes accessible and can influence person judgements[47,48]. This semantic form of person knowledge underpins major early theories of implicit social cognition[16,49–52] and intergroup bias[53–55], as well as more contemporary perspectives on intersectional and multidimensional impression formation[56–62].

It is unsurprising that theories of impression formation are dominated by semantic models. Because semantic information is declarative, it is very salient in the mind of a social perceiver[63]. Moreover, semantic knowledge is highly functional in a complex social milieu because it affords precision, nuance and flexibility; complex semantic impressions draw from a rich descriptive lexicon and can therefore describe a person from multiple angles and across contexts[64,65].

Updating of semantic knowledge occurs not through change in existing knowledge, but rather through elaboration based on new learning[66]. In the context of impressions, a perceiver might learn new complementary or contradictory trait information about a person — for example, that a professor who was strict in the classroom is supportive and fun in the lab. Existing knowledge can also be reinterpreted in light of new information[67]: after discovering that the professor is supportive, their strictness in the classroom is reinterpreted as encouragement. Although infrequently activated aspects of an impression can be forgotten over time through retrieval-induced forgetting[68,69], old trait information is typically retained alongside new knowledge, and a perceiver can select relevant new information when forming explicit judgements or summary evaluations[11,70]. Thus, while a perceiver's impression of Bob, the junk-hoarding neighbour, improves when they learn that he recycles toys for sick children, the knowledge of Bob as a hoarder remains.

Episodic memory

Episodic memory encodes multimodal snapshots of discrete experiences, from the extraordinary, such as the moment in the delivery room

when a parent first sets eyes on their newborn child, to the mundane, such as yesterday's lunch transaction at the local deli[71]. Early evidence that episodic memory functions as an independent system came from studies of people with brain lesions. In a famous case, the removal of patient H.M.'s medial temporal lobe (including the hippocampus) to treat his severe epilepsy left him unable to form new episodic memories. Yet, he retained knowledge of facts and the ability to play the piano — capacities that rely on semantic and instrumental memory[25]. Since then, neuroimaging studies of the medial temporal lobe in people with brain lesions and in healthy individuals have further established episodic memory as a separable memory system[35,72].

In impression formation, episodic memory supports the multisensory encoding of an event's details: the smell of autumn air, a friend's well-rehearsed words, their fiancé's surprised look, the sparkle of a diamond and the cheer of a crowd. Episodic memories provide specific examples of a person's behaviour that can inform an impression[73–75]. For example, episodic recall of how much a person shared in a prior interaction relies on the hippocampus[76] and informs a perceiver's choice of whether or not to engage with that person again[77].

Episodic memory can also provide a basis for semantic inference[78,79]. A perceiver can infer trait characteristics from episodes of a person's behaviour deliberatively through attribution (for example, reasoning that Sarah is generous because she leaves large tips)[3] or automatically through spontaneous trait inferences (for example, encoding Sarah as 'generous' without conscious deliberation)[7]. Similarly, a discrete episodic memory (for example, vividly recalling 100 people at an event) can give rise to a 'gist' memory ('there was a big crowd'). Although episodic memory and gist memory are simultaneously encoded[80], episodes guide specific judgements ('we will need 100 chairs') whereas gist guides more general, flexible judgements ('we will need more chairs than usual')[81].

Episodic memory updating is not incremental, in that it does not change gradually along a single trait or evaluative dimension. Similar to semantic memory, new episodes are typically encoded alongside older episodes; for example, a perceiver can remember separate instances when their friend was either outgoing or reclusive. However, one's memory for an episode can change through reconsolidation: when an episode is retrieved, it can be modified or integrated with new information and then re-encoded in its modified form[82–84]. Episodic memories can also be distorted through the imagination of past or future events, which might produce misremembering[85,86]. When misremembering occurs, it often does so in a way that supports one's motives[87,88] or is consistent with a stereotype[89–91] or a schema for how particular traits should align[92]. Changes such as these to episodes in memory, which might form the basis of an impression, can thereby contribute to impression updating.

Instrumental learning

Social interactions involve the exchange of actions and feedback — a perceiver gestures, the partner nods — and through this process, repeated throughout an interaction, the perceiver forms a behaviour-based preference toward the partner. This process involves instrumental learning (also known as operant conditioning), an action-based form of learning in which behaviours are associated with outcomes through reinforcement[93]. In contrast to semantic and episodic memories, instrumental learning is encoded in terms of reward value via activity in the striatum and expressed directly in behaviour[94–96]. Instrumental learning encompasses both goal-directed learning, which supports intentional, reward-driven behaviour, and habits, which support automatically cued responses[97,98].

Goal-directed instrumental learning. In goal-directed instrumental learning, the reward value of an action (such as approaching an object or person) is learned through choice and feedback. Following rules of reward reinforcement learning, choices that result in positive feedback are repeated and those that result in negative feedback are not[99]. Instrumental learning occurs incrementally in response to reward prediction errors (the discrepancy between expected and experienced reward feedback), such that reward associations develop across repeated experiences with action and feedback[100]. Because instrumental learning involves the encoding and expression of preferences through action, it has been theorized to underlie the behavioural (or conative) component of attitudes[20,21] and the priming of goal-directed behaviour[101].

Instrumental learning is further distinguished by its nondeclarative (or implicit) operation, such that instrumental learning associations can be formed and expressed without deliberation or apparent awareness[102,103]. For example, on probabilistic reinforcement tasks that involve incremental learning and therefore require the ability to track accumulated feedback across many trials, people with amnesia (who lack hippocampal function but retain normal striatal function) learn to make correct behavioural choices but appear unaware of what they have learned[104]. In healthy individuals, nondeclarative instrumental learning is often expressed as a skill — a well-practised, goal-directed action sequence that proceeds with little thought, such as playing the piano, driving a standard-transmission car or swinging a golf club[105].

In the context of impression formation, instrumental learning governs how people learn about others through direct social interaction — that is, through the exchange of action and feedback with another person[21,106,107]. Research that used functional magnetic resonance imaging in combination with computational modelling showed that when forming an impression through direct interactions, a perceiver encodes the instrumental reward value of approaching a partner in addition to inferring the partner's trait characteristics, and that these representations jointly contribute to decisions about the partner[106]. Instrumental associations have been likened to a gut feeling or intuition[108]; in the context of person impressions, they function implicitly to guide social choices independently of explicit traits or attitudes[64,109–112].

In contrast to semantic and episodic updating, instrumental associations are updated incrementally in response to prediction errors to maintain a running representation of a reward-based preference (that is, expected value)[99,113]. A prediction error occurs when feedback is more positive or negative than expected. The degree of updating in response to feedback depends on the size of the prediction error as well as the learning rate (that is, the weighting of new information), resulting in a revised expected value. This form of incremental preference updating resembles the kind of gradual evaluative change that is often examined in studies of impression updating[114].

Habits. Frequently enacted behaviours, whether goal-directed or not, can transform into a habit — a behaviour that is automatically triggered in response to an associated cue despite being contradictory or irrelevant to one's goals[115]. Whereas goal-directed instrumental learning is associated with reward processing in the ventral striatum, habits are associated with dorsal striatum activity[97,98,116].

In social contexts, habits are expressed when a person's presence activates an automatic behavioural response — for example, reflexively flashing a smile to a passing colleague or holding the door for a friend[21,109,117]. Such habits are often adaptive: they can enhance the fluency of social interactions by requiring less deliberation, and an impression based in habit might be more resistant to inconsistencies

in a partner's behaviour. However, habits might become maladaptive when a partner or relationship changes: a person might offer a beer to a friend who has quit drinking or mindlessly text an old partner post-breakup. In either case, a habit's unresponsiveness to feedback makes it extremely resistant to change. Indeed, evidence that instrumentally learned person preferences persist after they are no longer goal-consistent supports the role of habit in impressions[106,109,111,118].

Pavlovian learning

Swiss neurologist Édouard Claparède famously described a patient with severe amnesia who greeted him each day as if they had never met. As the story goes, one day the doctor held a tack in his hand, which pricked the patient during their handshake. The next day, despite again having no recollection of the doctor, the patient hesitated before shaking his hand — apparent evidence of fear learning without awareness of its cause. This classic account suggested a unique effect of Pavlovian fear learning on social impressions.

Pavlovian learning (also known as classical conditioning) refers to both a method and a mechanism; its mechanism describes how a neutral stimulus (conditioned stimulus) comes to evoke a response through its pairing with an aversive or appetitive stimulus (unconditioned stimulus)[24]. Pavlovian learning can be aversive or appetitive, although most research has focused on aversive (fear) conditioning[119]. Importantly for the present purposes, Pavlovian learning represents a stimulus–outcome association — that is, a reaction to a stimulus such as arousal or freezing — as opposed to a conceptual or action association, and it is further distinguished from other memory systems by its substrate in amygdala-related circuitry[120,121]. Pavlovian aversive conditioning functions nondeclaratively[35,122,123] and is expressed primarily as behavioural freezing, autonomic arousal and heightened attentional vigilance[124]. In the context of impression formation, a Pavlovian aversive component would represent a threat-based association, distinct from mere dislike or other negatively valenced responses[125,126].

In attitudes and impressions research, the term 'classical conditioning' is sometimes invoked to describe evaluative conditioning[121,127–129]. However, evaluative conditioning procedures typically involve the pairing of conceptual stimuli (for example, words or complex pictures, as opposed to an electric shock or food reward) and are measured as expression of conceptual valence judgements or associations, rather than physiological arousal or behavioural freezing, and therefore the effects of evaluative conditioning are more consistent with a semantic learning mechanism than with an amygdala-mediated Pavlovian learning mechanism[126,127,130,131].

Pavlovian-conditioned associations are not directly updated; instead, new associations are formed alongside existing associations[132]. These new associations can inhibit the expression of older threat associations (in aversive conditioning) to produce extinction; however, because the original associations remain, learned fear is easily re-established. It might be possible to change Pavlovian associations through reactivation and reconsolidation[133–135], but it remains unclear whether this intervention changes the underlying association or only its expression in behaviour[136,137].

Although aversive conditioning is robust in humans[138] and has been proposed as a component of intergroup bias[125,126,139,140], its role in impression formation has not been systematically investigated. Nevertheless, many existing findings are consistent with a role for Pavlovian fear conditioning in social impressions. Research on implicit prejudice is consistent with a distinction between associations rooted in Pavlovian threat and semantic valence, with prejudiced behaviour more strongly associated with threat-based associations[141]. In studies of intergroup interaction, a Pavlovian form of prejudice is consistent with evidence of social distance, stilted speech and action, interaction anxiety and fear-related affect — much like Claparède's famous patient[142–146]. Although more research is needed to determine the role of Pavlovian learning in impression formation, these findings suggest that it supports affective and threat-related behavioural responses to persons and groups.

In sum, a memory systems analysis clarifies that people learn about and represent persons through multiple learning mechanisms: semantic, episodic, instrumental and Pavlovian. These memory systems are separable, with unique operating characteristics and distinct neural substrates, and they function to produce specific kinds of social behaviour (Fig. 2). The multilevel person representation they create constitutes a holistic impression — a collection of knowledge, beliefs, preferences and opinions about a person, as well as the affective reactions and approach tendencies of the perceiver toward the person. Although impressions have long been considered multifaceted[1,2,147], this analysis specifies the mechanisms that support these facets and their unique roles in social behaviour.

Independent and interactive effects

Despite their unique features, learning and memory systems typically function in concert: during impression formation, perceivers simultaneously encode episodic information about the event, infer semantic knowledge about the person's traits and characteristics, develop a behavioural disposition through instrumental feedback, and form affective associations through Pavlovian learning[21]. Learning and memory systems also interact, whereby one memory system shapes or competes with another[148,149]. A key advance provided by a learning and memory analysis is that different aspects of an impression — subserved by semantic, episodic, instrumental or Pavlovian systems — are expressed in different ways, and that a consideration of their independent and interactive effects is essential for predicting how person impressions guide behaviour.

Independent effects on person impressions

Independent effects refer to cases where two or more memory systems have simultaneous yet unique effects on judgement or behaviour. Here, I describe such effects as they relate to implicit evaluation and the interplay of traits and evaluations.

Multiple forms of implicit evaluation. Implicit evaluation refers to the indirect expression of positive or negative evaluation towards a person or object[150]. The construct of implicit evaluation — that is, how implicit evaluations are formed, represented in the mind and expressed in behaviour, and whether they function automatically or unconsciously — is central to impression-formation research[37], yet it has been difficult to explain[151,152].

From a learning perspective, implicit evaluation reflects the operation of one or more different underlying memory systems[21,130]. For example, implicit evaluation could reflect an instrumental reward or Pavlovian threat association, both of which operate nondeclaratively and are expressed implicitly. Alternatively, implicit evaluation could reflect semantic knowledge which, although declarative and therefore subject to awareness, can be expressed indirectly on implicit tasks. In many cases, an implicit evaluation involves a combination of these systems. Considering the memory system basis of an implicit evaluation clarifies its features, function and expression and its potential for change.

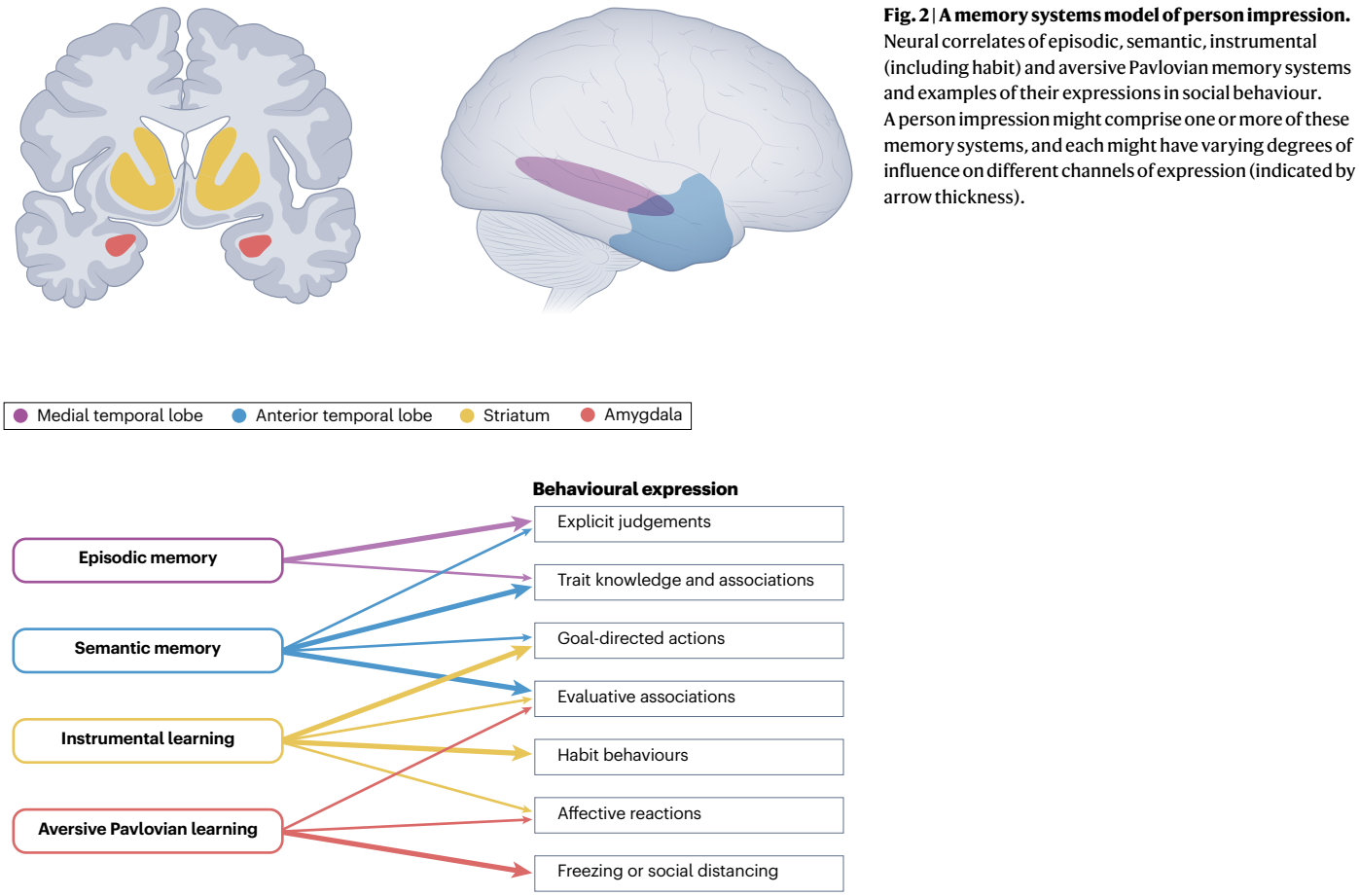

Fig. 2 | A memory systems model of person impression. Neural correlates of episodic, semantic, instrumental (including habit) and aversive Pavlovian memory systems and examples of their expressions in social behaviour. A person impression might comprise one or more of these memory systems, and each might have varying degrees of influence on different channels of expression (indicated by arrow thickness).

Nearly all existing studies of implicit evaluation concern semantic memory because they rely on tasks that primarily assess semantic associations between concepts and categories, such as evaluative priming tasks[153], the implicit-association test[154], and the affect-misattribution procedure[155]. Although some early models of implicit evaluation proposed a basis in affect[50,153,156], evidence for these accounts relied on data from semantic-categorization tasks which, in subsequent work, were shown to assess semantic rather than affective associations[157,158]. Thus, conventional implicit-evaluation tasks, which rely on semantic categorization, are now understood to be primarily sensitive to semantic and not affective associations[130,131,159–161]. As such, they can further be understood as reflecting knowledge that is declarative but observed indirectly when assessed with an implicit task.

An affective form of implicit evaluation has been proposed to correspond to a Pavlovian association[21,125,126], which can be assessed by physiological measures of skin conductance (an autonomic arousal indicator of either positive or negative affect depending on the elicitor) or the startle eyeblink response (an index of amygdala activity associated with the Pavlovian threat response)[162]. In early research on impression formation, a larger skin conductance response, suggesting implicit affective arousal, was associated with greater attraction towards agreeable partners[163]. In the intergroup domain, the startle eyeblink method was used to assess White American participants' implicit affective responses to Black, White and Asian faces[125]. This research found that the startle response was larger when participants viewed Black faces, relative to White or Asian faces, revealing a negative affective association that could not be explained by semantic processing. These studies identify an affective form of implicit evaluation rooted in Pavlovian learning that functions nondeclaratively, is expressed in physiological arousal and defensive behaviours, and is distinct from implicit evaluations based in semantic memory.

Implicit evaluation can also be represented by instrumental reward associations. Research using probabilistic reinforcement-learning tasks in which participants choose to interact with individuals and receive reward or non-reward feedback shows that individuals form preferences for people through instrumental learning[64,106,110–112,164]. These instrumental preferences (expressed in choice behaviour) predict subsequent social decisions independently of self-reported preferences and implicit association test measures of implicit evaluation[111,112]. Consistent with models of instrumental learning, this form of implicit evaluation operates implicitly and is expressed most directly in goal-directed behaviour — features that align it with motivation-oriented theories of social cognition[165,166].

Together, these findings clarify that 'implicit evaluation' can refer to different underlying memory processes — semantic, affective (Pavlovian), instrumental or some combination — and that a consideration of underlying memory process informs how an evaluation is formed and expressed. This analysis also highlights that appropriate measures are

needed to observe different forms of evaluative association (Fig. 3), and that theories of implicit evaluation built only on models of semantic memory and data from conventional implicit tasks are incomplete.

Traits versus evaluations. Traits and evaluations have long been distinguished in both impression formation and intergroup bias[1,2,53,167–169]. Traits, like stereotypes, refer to person or group characteristics and are represented as beliefs and conceptual associations in semantic memory. By contrast, evaluations refer to a perceiver's preference for an individual or group and, as described above, could reflect semantic, Pavlovian and instrumental associations.

In the intergroup domain, stereotypes (traits) and prejudice (evaluations) are difficult to discern because group stereotypes are often positive or negative in valence. However, research that used measures designed to unconfound stereotyping and evaluation — for example, such that measures of prejudice did not include stereotypes and measures of stereotyping were equated on valence — has observed weak correlations between stereotypes and prejudice[43,156,161,170–172]. Research on spontaneous impression formation has similarly found dissociations in the formation and effects of trait and evaluative inferences[19,173].

Notably, this trait–evaluation distinction is at odds with the position that stereotypes and prejudice emerge from a single underlying representation[174]. However, evidence for the single-representation position has come from measures or manipulations that confound stereotype traits with valence[174,175]. For example, one study found large correlations between implicit-association test measures of implicit prejudice and stereotyping when stereotypes with positive and negative valence were used[174]. However, when unconfounded implicit-association test measures of prejudice and stereotyping were used, the intercorrelations were relatively small, suggesting separate processes, and similar in effect size to prior findings, supporting a stereotype–evaluation distinction[43,156]. Thus, these findings further support the idea that although trait associations and evaluations are often related, they reflect different underlying representations.

Independent effects on behavioural expression

A key contribution of a memory systems analysis of social cognition is that it predicts how impressions are expressed in behaviour (Fig. 2). Whereas semantic and episodic impressions guide explicit thoughts, judgements and plans regarding a person, instrumental associations implicitly guide actions during decision-making and social interactions. Habits guide automatic actions to previously rewarded cues, whereas Pavlovian associations prepare an organism to respond to potential threats through freezing, attentional vigilance and physiological readiness[176].

An early demonstration of these effects showed that White Americans' scores on an implicit-association test measure of implicit stereotypes uniquely predicted their trait impressions of a Black partner, whereas scores on an implicit prejudice implicit-association test (proposed at the time to reflect an affective Pavlovian association) uniquely predicted their seating distance from a Black partner[156]. Other research showed that feelings of intergroup anxiety (associated with a Pavlovian response) selectively increased the expression of implicit prejudice but not implicit stereotypes[170]. These patterns resemble previous dissociations between explicit cognitive and affective measures of intergroup bias[169] and between effects of explicit prejudice beliefs and implicit race evaluations on behaviour in interracial interactions[142,143,177].

Research has also distinguished the effects of trait-based and reward-based impressions on participants' social decisions[106]. Whereas instrumental reward associations tend to be more strongly expressed in behavioural choices to interact with partners, semantic trait associations are more strongly expressed in self-reported social preferences and intentions for future interaction[106,110,112]. In other work, impressions based in episodic knowledge (recalling the exact amount a person had donated to a charity) and semantic knowledge (a gist description of the donation as 'some' or 'none') were shown to play different roles in decisions to help someone[81]. The dissociation between semantic and episodic aspects of an impression has also been demonstrated using a directed forgetting procedure: although instructions to forget a behaviour associated with a person impaired later episodic memory for the behaviour, the trait implied by the behaviour remained semantically accessible and continued to influence person judgement[178].

In cases where two or more memory systems compete to influence a response, the expression of one over another might be moderated by situational factors. For example, although episodic and instrumental learning normally function in concert, cognitive load selectively impairs episodic memory, leaving instrumentally learned responses intact to solely guide performance[179]. A similar pattern has been shown in the context of impression formation: although perceivers formed spontaneous trait and evaluative inferences simultaneously, cognitive load selectively impaired the expression of trait inferences but not evaluative inferences[19].

The timing and certainty of information during learning can also affect the expression of competing memory systems. Studies of feedback-based learning show that people simultaneously form episodic and instrumental associations when feedback is immediate, but instrumental learning is selectively impaired and only episodic learning occurs when feedback is delayed by even a few seconds[180,181]. Similarly, in uncertain environments, people rely more on episodic than on instrumental learning in decision making, consistent with a shift from automatic to deliberative processing[182,183]. These findings have implications for how impressions are expressed in situations marked by feedback delay or uncertainty, such as in online communication.

In sum, different components of an impression (semantic, episodic, instrumental and Pavlovian) are expressed via different response channels, and their expression can be moderated in specific ways by situational factors.

Interactive effects

Memory systems also function interactively, such that one can shape another's operation[149]. Such interactions have been demonstrated extensively in nonsocial domains[148,179,184], and they are likely to have similar effects in social contexts[21,185].

A well-known example of memory system interaction is that Pavlovian fear enhances the activation and consolidation of episodic memory, reflecting the influence of amygdala activity on hippocampal function[186–188]. This Pavlovian–episodic interaction suggests that fear-based arousal in particular should enhance the encoding of episodic person memory — a prediction consistent with observations of a negativity bias in impression formation, in which negative information weighs more heavily in impressions[189].

An interactive effect of semantic and instrumental systems has also been examined in prejudice formation[164]. This research found that mere knowledge of a societal stereotype (a form of semantic memory) implicitly shapes how a perceiver experiences and learns from members of the stereotyped group through its effect on instrumental learning in subsequent social interactions, leading to the internalization of prejudice. This form of memory system interaction, between

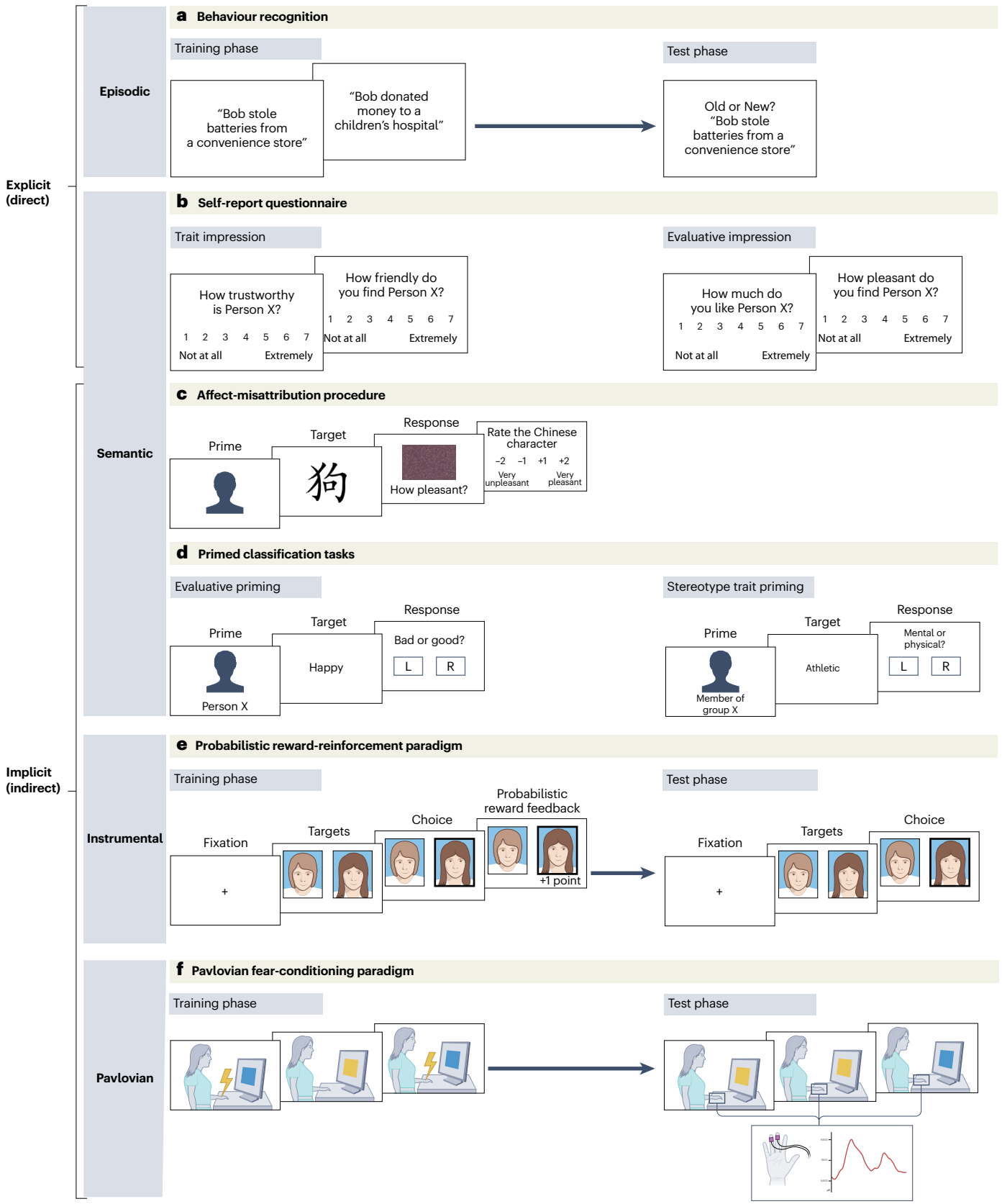
Explicit (direct)
Implicit (indirect)
Episodic
Semantic
Instrumental
Pavlovian
a Behaviour recognition
Training phase
"Bob stole batteries from a convenience store"
"Bob donated money to a children's hospital"
Test phase
Old or New? "Bob stole batteries from a convenience store"
b Self-report questionnaire
Trait impression
How trustworthy is Person X?
1 2 3 4 5 6 7
Not at all
Extremely
How friendly do you find Person X?
1 2 3 4 5 6 7
Not at all
Extremely
Evaluative impression
How much do you like Person X?
1 2 3 4 5 6 7
Not at all
Extremely
How pleasant do you find Person X?
1 2 3 4 5 6 7
Not at all
Extremely
c Affect-misattribution procedure
Prime
Target
狗
Response
How pleasant?
Rate the Chinese character
−2 −1 +1 +2
Very unpleasant
Very pleasant
d Primed classification tasks
Evaluative priming
Prime
Person X
Target
Happy
Response
Bad or good?
L
R
Stereotype trait priming
Prime
Member of group X
Target
Athletic
Response
Mental or physical?
L
R
e Probabilistic reward-reinforcement paradigm
Training phase
Fixation
+
Targets
Choice
Probabilistic reward feedback
+1 point
Test phase
Fixation
+
Targets
Choice
f Pavlovian fear-conditioning paradigm
Training phase
Test phase

Fig. 3 | Experimental paradigms for assessing impression formation. Explicit (direct) and implicit (indirect) assessments of impressions tap into different learning and memory mechanisms. **a**, Episodic memory can be assessed using a behaviour-recognition task, in which participants read a series of behavioural descriptions in a training phrase and then, in a subsequent test phase, indicate whether behavioural descriptions are old or new. **b**–**d**, Semantic memory can be assessed with self-report questionnaires on which participants rate a target's attributes on a Likert-type scale (**b**), primed self-report tasks (such as the affect-misattribution paradigm) in which participants rate their evaluation of the target stimulus that follows presentation of an attitude object prime (**c**) or primed conceptual-classification tasks in which participants classify the conceptual or valence category of target words that follow the presentation of an attitude object prime (**d**). **e**, Instrumental learning can be measured using probabilistic reward-reinforcement paradigms, in which participants choose (through button press actions) between two targets and receive probabilistic reward feedback based on their choice. Learning of reward associations is assessed in a subsequent test phase, in which participants choose between targets without feedback. **f**, In a version of a Pavlovian fear-conditioning paradigm, participants are conditioned by receiving a shock every time a conditioned stimulus is presented. During the test phase, in which no shocks are administered, learning is indicated by autonomic arousal in response to the conditioned stimulus as measured by skin conductance response.

declarative semantic knowledge and nondeclarative instrumental learning, describes a process through which exposure to societal-level social information can induce individual-level implicit attitudes[112,164].

Semantic knowledge (such as a pre-existing preference or stereotype) can also prevent an individual from engaging in instrumental social-interactive learning. For example, if a person holds a positive impression of a particular group, they might selectively interact with its members and thereby never form or update impressions of other groups[190,191]. This selective exposure effect has been proposed as a mechanism through which group prejudices and stereotypes are formed and maintained[190,192,193].

In sum, although research has just begun to explore interactive memory system effects in impression formation, this approach promises to advance our understanding of how impression components such as traits, stereotypes and evaluations are formed and expressed — often implicitly — in different social contexts.

Impression updating

Once an impression is formed, it can continue to change through impression updating. Although Will Rogers famously quipped, "You never get a second chance to make a first impression," this depends on how the impression was formed: whereas instrumental associations and semantic knowledge are readily revised, changes in episodic memory and Pavlovian associations are not.

Much research on impression updating examines changes in evaluation — that is, how new trait information about a person incrementally changes the positivity or negativity of an impression[1,194]. This focus on evaluative updating (as opposed to trait updating) might reflect the specific mechanisms through which information is updated in different memory systems. As described above, trait concepts are represented in semantic memory, which is not updated in an incremental fashion but instead incorporates new trait knowledge. By contrast, evaluations can be supported by semantic, instrumental or Pavlovian memory processes; of these, only instrumental associations are updated incrementally. Thus, conceptualizations of incremental impression updating align most closely with an instrumental learning mechanism, whereas categorical changes, such as revisions of trait concepts or reversals in evaluative concepts, are more consistent with a semantic memory mechanism.

Few studies to date have directly examined the implications of memory systems for impression updating. In one relevant programme, distinct patterns of trait and evaluative updating were found in the context of spontaneous trait and evaluative inferences. Prior findings showed that spontaneous trait and evaluative inferences comprise distinct representations, formed in parallel[19]. Building on this work, another study found that only spontaneous evaluative inferences were updated in response to new impression-inconsistent information about a target's behaviour, consistent with an instrumental learning process; by contrast, the updating of spontaneous trait inferences involved the encoding of new traits alongside the old traits[173], consistent with a basis in semantic memory. Although both new and old traits remain in memory, a perceiver can selectively base their predictions for a person's future behaviours on the new trait information[195]. Thus, a consideration of memory system function clarifies how different forms of spontaneous inference are updated in memory and applied selectively to social judgements.

Research on the instrumental learning of impressions has used computational reinforcement learning models to demonstrate updating[196,197]. Consistent with reinforcement learning theory[99], these models specify the incremental, trial-by-trial updating of a reward association (expected value) in response to new information. By showing that behavioural data from instrumental impression-formation tasks fit best to such models, these studies provide strong evidence for an instrumental learning mechanism of updating[106,112,164].

Given the different expressions of memory systems in behaviour, an assessment of updating must be sensitive to the underlying representation of interest. Measures that rely on self-report, including questionnaires and some implicit tasks such as the affect-misattribution procedure, are primarily sensitive to changes in semantic associations. Measures that rely on action (for example, behavioural classifications or choices) are more sensitive to changes in instrumental associations. Behavioural tasks that pick up on freezing or response slowing and physiological measures of autonomic arousal are more sensitive to changes in Pavlovian threat associations. A task might be sensitive to multiple underlying memory processes to the extent that it combines these response features (such as the implicit-association test, evaluative priming and some versions of the affect-misattribution procedure). Furthermore, when there is a mismatch between the measure and the underlying learning process, updating effects might be obscured.

A consideration of mechanism–measure match can illuminate longstanding questions about the nature of impression updating, such as whether implicit impression updating occurs slowly[114] or rapidly[194]. In one set of experiments[114], participants formed impressions of a target person by reading statements about a behaviour, deciding whether it was true of the target, and then receiving feedback on whether their choice was correct — a task that involves elements of both semantic and instrumental learning. In these studies, a change in the valence of target behaviours produced a rapid change in evaluation on a self-report measure but a relatively slow change on the implicit-association test, an implicit task that involves behavioural choice classifications. By contrast, studies that used a similar impression-formation task found

Table 1 | Implicit and explicit components of person impressions and attitudes

	Learning and memory system	Role in person impressions
Explicit (declarative) processes	Semantic	Knowledge of a person's traits, goals, attitudes or circumstances Knowledge of one's own beliefs and preferences regarding a person
	Episodic	Multimodal recollection of events associated with a person
Implicit (nondeclarative) processes	Instrumental	Goal-directed behavioural tendencies towards a person learned via direct interaction with them Habitual response tendencies towards a person
	Pavlovian	Physiological affective response to a person

that exposure to a single extreme countervailing behaviour by the target person produced rapid updating on both explicit and implicit evaluation measures[194,198]. Notably, in these latter studies implicit evaluation was measured using the affect-misattribution procedure, a task in which participants make evaluative self-report judgements of targets following a positive or negative prime[155]. Thus, it is possible that the discrepancy between findings in part reflects the different implicit measures: whereas responses on both the affect-misattribution procedure and implicit-association test involve a combination of semantic and instrumental processes, the affect-misattribution procedure should be more sensitive to semantic knowledge than the implicit-association test, and should therefore reveal more rapid updating. Indeed, when the affect misattribution was replaced with an implicit-association test measure of updating in studies examining updating in response to an extreme countervailing behaviour, the signature reversal in impression valence was not observed[199]. Thus, apparent discrepancies in the updating of implicit evaluations might reflect different underlying forms of memory (semantic or instrumental) and the use of measures that are differentially sensitive to these underlying memory systems.

In sum, a memory systems framework distinguishes patterns of impression updating associated with different underlying learning processes and clarifies how each can be measured. Furthermore, a memory systems analysis can be helpful for interpreting existing patterns of impression updating, and it provides a basis for developing interventions for impression change.

Implications for current debates

Social cognition researchers have long debated the meaning of implicit impressions and attitudes — for example, whether they can operate nonconsciously — and, relatedly, whether impressions and attitudes represent single, dual or multiple underlying processes. A memory systems analysis illuminates these debates by considering contemporary memory research that extends beyond conventional models of social cognition.

The nature of implicit impressions

Few topics in social cognition spark as much debate as the nature and utility of implicit processes and the tasks designed to measure them. Since their emergence, researchers have debated whether implicit processes are truly nonconscious or merely indirect, and whether (and how) these processes relate to behaviour. While many scholars have called for greater clarity in defining the construct[151,200], some suggest abandoning it altogether[152]. However, from a learning and memory perspective, these debates stem largely from the limitations of social cognition theories that narrowly assume a basis in semantic memory.

The memory systems literature offers a more nuanced understanding of implicit processes, in part because it incorporates studies of human cognition with studies of nonhuman animals[35] (such as rats[28,31,120] or aplysia (sea slugs)[201]) — subjects that cannot self-report and might lack the capacity for semantic cognition. These studies required the development of models of learning and behaviour — such as Pavlovian conditioning and instrumental learning — that do not rely on explicit reports or semantic processes and which appear to operate implicitly alongside semantic and episodic memory in humans. Furthermore, research using animal models permits the identification of neural circuits underlying these implicit forms of learning and behaviour, enabling these processes to be theoretically distinguished from other cognitive mechanisms.

Studies of people with brain lesions have particularly elucidated the nature of implicit memory processes. Research on people with damage to the temporal lobe (such as H.M.) demonstrates that implicit associations that involve Pavlovian or instrumental learning can occur without declarative knowledge of what was learned. Conversely, people with amygdala damage or Parkinson's disease can learn explicit associations based on semantic or episodic memory in the absence of Pavlovian or instrumental learning, respectively[102,122,187]. Moreover, neuroimaging studies of healthy individuals show that although the neural substrates of implicit and explicit processes are dissociable, they frequently co-occur and might create the appearance of a unified response in behaviour[179]. Findings such as these reveal that implicit and explicit processes involve the coordinated activity of multiple memory systems rather than a single (for example, propositional) mechanism.

This body of evidence has inspired a range of experimental tasks and methods designed to isolate multiple forms of learning and memory, including Pavlovian, instrumental, semantic, episodic or combinations thereof. This approach can be contrasted with the use of tasks in social cognition research that assume a basis in only semantic processing — a constraint that limits the measurement and interpretation of implicit or nonconscious processes.

This approach to measurement in learning and memory research, in which tasks are tailored to the way a particular underlying memory system is expressed, informs the assessment of implicit social impressions. Some forms of memory — episodic and semantic — are declarative (reportable), can be expressed directly (explicitly) and are typically subject to awareness (Table 1). Thus, semantic and episodic associations can be assessed directly using explicit measures (such as self-reports), yet they can also be observed indirectly in implicit tasks that assess conceptual associations (such as semantic priming). By contrast, nondeclarative forms of memory — instrumental (including habits) and Pavlovian — are expressed indirectly and can operate outside conscious awareness. As such, instrumental and Pavlovian associations can be observed using indirect (implicit) measures (such as probabilistic classification or fear-conditioning tasks), because these associations are not directly accessible to awareness and are therefore not reportable.

It is notable that most implicit social cognition tasks (such as the implicit-association test, affect-misattribution procedure, and semantic or evaluative priming tasks) blur the distinctions between memory systems. That is, they measure semantic associations (of traits

or evaluations) with an indirect assessment. Although such tasks might give the appearance of a nonconscious semantic association, a learning and memory analysis suggests that such measures capture the indirect expression of declarative (conscious) knowledge. This interpretation is consistent with evidence suggesting that people are often aware of associations expressed in implicit tasks, even if the expression is unintentional[202–205].

To measure nonconscious associations, methods tailored to non-declarative memory processes are required. For instance, instrumental learning can be assessed using tasks that afford the formation of action–reward associations while hindering semantic learning, such as probabilistic selection tasks[102,206]. Studies of impression formation have adapted these tasks to demonstrate implicit social preferences that are independent of the subjective attitudes or semantic associations of participants[106,112,164]. Similarly, Pavlovian learning can be assessed using Pavlovian fear-conditioning paradigms paired with measures of freezing or autonomic arousal[207]. Habits can be assessed using reward-devaluation tasks, which measure learned behaviours that persist after they are no longer goal-relevant[97]. Critically, some tasks engage a combination of memory processes, whereas others might assess only one component of a multi-system response. Careful task design and interpretation are therefore crucial to isolating and understanding impression representations of interest.

In some instances, a perceiver might be aware of possessing a belief or association but unaware of its expression in behaviour or the processes through which it is expressed[151]. For example, a mathematics professor might be aware of their gender stereotype beliefs but unaware of how these beliefs influence their grading decisions. Similarly, a participant in an impression study might be aware of their stereotype knowledge but unaware of how it produces bias on an implicit task. This phenomenon can be explained by the interplay between semantic and instrumental processes in most implicit tasks; that is, although one's belief is represented in declarative semantic memory, the influence of this belief on task behaviour (which involves target classifications in semantic priming) relies on a non-declarative instrumental process — an interplay between semantic and instrumental systems that is neither consciously accessible nor easily controlled[164,208]. Thus, a memory systems analysis helps to clarify which kinds of associations are implicit and how to measure them, and why some aspects of an implicit task are subject to awareness whereas others are not.

Single- versus dual- versus multi-process accounts

Another longstanding debate concerns the number of processes needed to explain impressions. If it is assumed that impressions are based on known mechanisms of learning and memory, then a multi-process account based on these memory systems is most plausible. Because this multi-process account is grounded in the functions and neural substrates of learning and memory, it offers a deductive, model-based approach to predictions about impression formation. The goal of this approach is not to determine the number of processes, but to understand their specific functions in social cognition[21,34].

By contrast, traditional dual- and single-process models reflect an inductive approach that attempts to explain the available data with the most parsimonious account. Dual-process models propose two general kinds of process: one that is associative, automatic, impulsive and nonconscious and one that is propositional, deliberative, reflective and conscious[50,52,166]. Although dual-process models vary in their particular aims and features, they generally explain divergent patterns of implicit and explicit responses as arising from these two types of processing. Single-process models posit that responses on both implicit and explicit tasks can be explained by a single propositional process — an account suggested by observations that single instances of explicit information can induce or change implicit evaluations[202,203,209] and that participants are often aware of associations assessed by implicit tasks[204,205].

From a learning and memory perspective, existing dual- and single-process models of social cognition can be viewed as concerning the operations of semantic memory — that is, the conceptual beliefs or associations measured with questionnaires and conventional implicit tasks. If a model's purpose were only to explain expressions of semantic memory, then a memory systems analysis aligns with a single-process propositional account. However, if the goal were to explain other forms of social behaviour, such as those guided by instrumental responses, habits, episodes or Pavlovian reactions, then neither single- nor dual-process accounts that assume knowledge representation in semantic memory alone are sufficient.

A key distinction between single-process and dual-process models is that dual-process models typically include a second process that modulates or supersedes an automatic semantic association. Although a memory systems framework primarily concerns the representation and expression of associations, it is similar to dual-process models in that it assumes that the expression of memory systems is modulated by cognitive control[63]. In this regard, a memory systems framework can be thought of as part of an expanded dual-process model that more precisely articulates both the representational and modulatory components of social cognition (see Box 2).

It is notable that while many existing dual-process models refer to associative networks that are semantic in nature, some influential models describe associative processes as affective or motivational — features that intuitively correspond to Pavlovian or instrumental processes[50,166]. Collectively, these existing models are suggestive of a multiple memory systems account. One implication is that such models, originally proposed within a dual-process framework, might benefit from re-specification within a memory systems framework. Doing so could improve their theoretical precision, expand their methodological repertoires (by using measures appropriate for the assessment of other memory systems) and increase their explanatory power.

Conclusions

The field of social cognition was borne of the insight that impression-formation processes are rooted in learning and memory, with early approaches importing theories of semantic memory from cognitive psychology to study person perception[29]. The current analysis continues this tradition by providing an updated perspective of impression formation informed by contemporary models of learning and memory. Here, I describe some key contributions of this updated approach.

First, a memory systems framework provides an expanded theoretical framework. That is, it broadens the scope of traditional impression-formation theories to include all the ways people experience and encode the social world, incorporating episodic, instrumental, Pavlovian, and habit components, a greater focus on behaviour, and a grounding in neural function. Moreover, it introduces the idea that different components of an impression can have interactive effects, and it provides a framework for how such interactions guide impression formation, expression and updating.

Second, a memory systems framework clarifies that conventional measures of social impressions and attitudes pertain primarily to semantic memory, and therefore their ability to assess aspects

Box 2 | Memory systems and cognitive control

Many theories of impression formation are dual-process models; that is, they posit separate processes for how person (or group) knowledge is represented in the mind — typically, in an associative semantic network — and how the use of this knowledge is modulated by cognitive control[5,6,49,52,53,215]. According to these models, control operates by inhibiting[215], adjusting[6] or replacing (with an alternative deliberative response)[53] automatically activated knowledge in order to promote an intentional response.

It is notable that the memory systems framework pertains mainly to the first (associative) process of most dual-process models — that is, the representation and expression of person knowledge. Yet, as in dual-process models, memory systems also interact with mechanisms of cognitive control. However, the way control is conceptualized in a memory systems framework differs from that of conventional dual-process models in two ways.

First, a memory systems framework acknowledges that learning and memory mechanisms vary in the degree to which they support more reflexive or intentional processes. Habit and Pavlovian associations are reflexive, and are therefore less aligned with intentions and more difficult to control. By contrast, although goal-directed instrumental associations are nondeclarative, they represent intentions and are therefore usually aligned with controlled processes. Semantic associations, which are declarative, are easy to control when expressed explicitly but difficult to control when expressed indirectly (that is, implicitly). That is, different kinds of association vary in their intentionality and controllability owing to differences inherent in the structure and function of their underlying memory systems.

Second, because a memory systems approach is part of a broader cognitive neuroscience framework, it naturally draws upon cognitive neuroscience models of control to understand how learned responses are regulated[21]. Whereas dual-process models have conceptualized control as operating directly on semantic activations in the mind[53,215], cognitive neuroscience models (which situate control in the prefrontal cortex and consider connectivity between this region and other structures in the brain) place greater emphasis on the regulation of behaviour (in addition to control effects on perception, attention and working memory)[216]. According to this model, control is engaged when conflict is detected between a behavioural tendency and an intended response[217] and then implemented by either inhibiting the unwanted action or selecting an alternative response[218,219].

This analysis of control suggests that strategies that target behaviour are more effective than strategies that target mental associations. This perspective aligns with evidence supporting the effectiveness of implementation intentions targeting behaviour[220] and behavioural correction strategies[221,222], as well as the ineffectiveness of mental suppression[223,224] for the control of unwanted social responses. In the context of impression formation, this model suggests that cognitive control most directly targets the behavioural expression of person knowledge rather than the activation of that knowledge itself.

The integration of a memory systems framework with this model of control suggests an expanded and reconceptualized version of a dual-process model. Although more complex than conventional dual-process models, this integrated model reflects contemporary knowledge regarding mechanisms of learning and memory, control, and behaviour relevant to social cognition. Importantly, research based on this model does not need to test all of its components at once. Rather, this model is useful for generating hypotheses regarding how specific aspects of an impression function (for example, the interplay of semantic and instrumental preferences) and are potentially controlled. By combining a memory systems framework with a control mechanism, researchers can understand better how impressions function across different situations.

of impressions that involve other forms of learning and memory is limited. It further suggests that the adoption of other methods from learning and memory research would permit access to a broader range of impression and attitude processes.

Third, a memory systems framework addresses existing theoretical debates by clarifying the role of awareness and implicit processes in impression formation, and contextualizing the single- versus dual-process debate within a broader set of learning and memory processes. It also elucidates the process of impression updating, accounting for both fast and slow modes of attitude change and explains why different components of an impression might be expressed in different kinds of response.

Finally, a longstanding critique of impression-formation research is that its measures often fail to predict behaviour. Whereas existing models typically focus on the formation, representation and activation of semantic impressions within the mind, a memory systems framework generates predictions for behaviour based on known neurocognitive pathways through which memory influences decision and action. As such, a memory systems framework provides an expanded account of how impressions are expressed in behaviour.

Having established a multiple memory systems basis for social cognition and attitudes, the next step is to develop and test predictions from this updated framework for uniquely social phenomena. Research has already begun to explore interactive effects of semantic and instrumental processes to understand how stereotypes influence impression formation in direct social interactions[112,164]. Other research has used this approach to examine the unique roles of episodic memory and habits in impression formation[81,109]. Although previous research has examined Pavlovian fear-conditioning effects in group-based impressions[125,140], questions on its interplay with other impression processes are ripe for exploration. As this approach develops, it can also be integrated with updated models of cognitive control and decision making to explain more fully how impressions are regulated to serve adaptive functions.

An important new direction in impression-formation research concerns the relation between individual-level impressions and societal-level factors. A memory systems framework aids this endeavour by specifying how individuals encode and internalize information from higher-level social structures and communicate it to others[164,210]. This approach provides a theoretical basis for situating impression formation in a multilevel framework that connects individual-level processes to cultural and systemic processes.

Progress towards these goals will require theoretical and methodological expertise that is increasingly interdisciplinary, incorporating, for example, theory and methods from cognitive neuroscience,

computational cognition, and sociology into the social psychology curriculum. At the same time, this increasingly interdisciplinary approach to impression formation highlights its utility as a hub domain within psychology for understanding the high-level functions of more basic cognitive processes. In doing so, this memory systems perspective brings the field closer to Asch's holistic conceptualization of impression formation[1] as a core capacity of the human mind.

Published online: 29 April 2025

References

1. Asch, S. E. Forming impressions of personality. *J. Abnorm. Soc. Psychol.* **41**, 258–290 (1946).
2. Carlston, D. E. in *Advances in Social Cognition* Vol. 7 *Associated Systems Theory* (ed. Wyer, R. S.) 1–78 (Erlbaum, 1994).
3. Jones, E. E. & Davis, K. E. in *Advances in Experimental Social Psychology* vol. 2, 219–266 (Academic, 1965).
4. Malle, B. F. & Holbrook, J. Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *J. Pers. Soc. Psychol.* **102**, 661–684 (2012).
5. Trope, Y. Identification and inferential processes in dispositional attribution. *Psychol. Rev.* **93**, 239–257 (1986).
6. Gilbert, D. T., Krull, D. S. & Pelham, B. W. Of thoughts unspoken: social inference and the self-regulation of behavior. *J. Pers. Soc. Psychol.* **55**, 685–694 (1988).
7. Winter, L. & Uleman, J. S. When are social judgments made? Evidence for the spontaneousness of trait inferences. *J. Pers. Soc. Psychol.* **47**, 237–252 (1984).
8. Hassin, R. R., Aarts, H. & Ferguson, M. J. Automatic goal inferences. *J. Exp. Soc. Psychol.* **41**, 129–140 (2005).
9. Heider, F. Social perception and phenomenal causality. *Psychol. Rev.* **51**, 358–374 (1944).
10. Read, S. J., Jones, D. K. & Miller, L. C. Traits as goal-based categories: the importance of goals in the coherence of dispositional categories. *J. Pers. Soc. Psychol.* **58**, 1048–1061 (1990).
11. Olcaysoy Okten, I. & Moskowitz, G. B. Goal versus trait explanations: causal attributions beyond the trait-situation dichotomy. *J. Pers. Soc. Psychol.* **114**, 211–229 (2018).
12. Ames, D. R. Inside the mind reader's tool kit: projection and stereotyping in mental state inference. *J. Pers. Soc. Psychol.* **87**, 340–353 (2004).
13. Kruse, F. & Degner, J. Spontaneous state inferences. *J. Pers. Soc. Psychol.* **121**, 774–791 (2021).
14. Brewer, M. B. in *A Dual Process Model of Impression Formation* (eds. Wyer, R. S. Jr & Srull, T. K.) 1–36 (Psychology Press, 1988).
15. Fiske, S. T. & Neuberg, S. L. in *Advances in Experimental Social Psychology* Vol. 23 (ed. Zanna, M. P.) 1–74 (Academic, 1990).
16. Kunda, Z. & Thagard, P. Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory. *Psychol. Rev.* **103**, 284–308 (1996).
17. Anderson, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *J. Exp. Psychol.* **70**, 394–400 (1965).
18. Cone, J., Mann, T. C. & Ferguson, M. J. in *Advances in Experimental Social Psychology* Vol. 56 (ed. Olson, J. M.) 131–199 (Academic, 2017).
19. Schneid, E. D., Carlston, D. E. & Skowronski, J. J. Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *J. Pers. Soc. Psychol.* **108**, 681–696 (2015).
20. Breckler, S. J. Empirical validation of affect, behavior, and cognition as distinct components of attitude. *J. Pers. Soc. Psychol.* **47**, 1191–1205 (1984).
21. Amodio, D. M. Social cognition 2.0: an interactive memory systems account. *Trends Cogn. Sci.* **23**, 21–33 (2019).
22. Collins, A. M. & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychol. Rev.* **82**, 407–428 (1975).
23. Hull, C. L. *Principles of Behavior: an Introduction to Behavior Theory* x, 422 (Appleton-Century, 1943).
24. Pavlov, I. P. *Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex* xv, 430 (Oxford Univ. Press, 1927).
25. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiat.* **20**, 11–21 (1957).
26. Shiffrin, R. M. & Schneider, W. Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* **84**, 127–190 (1977).
27. Skinner, B. F. *The Behavior of Organisms: An Experimental Analysis* (Skinner Foundation, 2019).
28. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
29. Hastie, R. et al. *Person Memory (PLE: Memory): The Cognitive Basis of Social Perception* (Psychology Press, 1980).
30. Gabrieli, J. D. E. Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* **49**, 87–115 (1998).
31. McDonald, R. J. & White, N. M. A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behav. Neurosci.* **107**, 3–22 (1993).
32. Squire, L. R. & Zola, S. M. Structure and function of declarative and nondeclarative memory systems. *Proc. Natl Acad. Sci. USA* **93**, 13515–13522 (1996).
33. Tulving, E. How many memory systems are there? *Am. Psychol.* **40**, 385–398 (1985).
34. Henke, K. A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* **11**, 523–532 (2010).
35. Squire, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* **82**, 171–177 (2004).
36. Cantor, N. & Mischel, W. Traits as prototypes: effects on recognition memory. *J. Pers. Soc. Psychol.* **35**, 38–48 (1977).
37. Uleman, J. S. & Kressel, Laura, M. in *Oxford Handbook of Social Cognition* (ed. Carlston, D. E.) 53–73 (Oxford Univ. Press, 2013).
38. Moskowitz, G. B. *Introduction to Social Cognition: The Essential Questions and Ideas* (Guilford, 2024).
39. Kumar, A. A. Semantic memory: a review of methods, models, and current challenges. *Psychon. Bull. Rev.* **28**, 40–80 (2021).
40. Binney, R. J. & Ramsey, R. Social semantics: the role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neurosci. Biobehav. Rev.* **112**, 28–38 (2020).
41. Olson, I. R., McCoy, D., Klobusicky, E. & Ross, L. A. Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc. Cogn. Affect. Neurosci.* **8**, 123–133 (2013).
42. Contreras, J. M., Banaji, M. R. & Mitchell, J. P. Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Soc. Cogn. Affect. Neurosci.* **7**, 764–770 (2012).
43. Gilbert, S. J., Swencionis, J. K. & Amodio, D. M. Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia* **50**, 3600–3611 (2012).
44. Wyer, R. S. The acquisition and use of social knowledge: basic postulates and representative research. *Pers. Soc. Psychol. Bull.* **6**, 558–573 (1980).
45. Carlston, D. E. & Skowronski, J. J. Savings in the relearning of trait information as evidence for spontaneous inference generation. *J. Pers. Soc. Psychol.* **66**, 840–856 (1994).
46. Srull, T. K. & Wyer, R. S. The role of category accessibility in the interpretation of information about persons: some determinants and implications. *J. Pers. Soc. Psychol.* **37**, 1660–1672 (1979).
47. Tory Higgins, E., Rholes, W. S. & Jones, C. R. Category accessibility and impression formation. *J. Exp. Soc. Psychol.* **13**, 141–154 (1977).
48. Bargh, J. A. & Pietromonaco, P. Automatic information processing and social perception: the influence of trait information presented outside of conscious awareness on impression formation. *J. Pers. Soc. Psychol.* **43**, 437–449 (1982).
49. Fazio, R. H. in *Advances in Experimental Social Psychology* Vol. 23 (ed. Zanna, M. P.) 75–109 (Academic, 1990).
50. Gawronski, B. & Bodenhausen, G. V. Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychol. Bull.* **132**, 692–731 (2006).
51. Greenwald, A. G. & Banaji, M. R. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* **102**, 4–27 (1995).
52. Smith, E. R. & DeCoster, J. Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Pers. Soc. Psychol. Rev.* **4**, 108–131 (2000).
53. Devine, P. G. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* **56**, 5–18 (1989).
54. Hamilton, D. L. & Sherman, S. J. Perceiving persons and groups. *Psychol. Rev.* **103**, 336–355 (1996).
55. Sherman, J. W. Development and mental representation of stereotypes. *J. Pers. Soc. Psychol.* **70**, 1126–1141 (1996).
56. Chen, J. M. An integrative review of impression formation processes for multiracial individuals. *Soc. Pers. Psychol. Compass* **13**, e12430 (2019).
57. Freeman, J. B. & Ambady, N. A dynamic interactive theory of person construal. *Psychol. Rev.* **118**, 247–279 (2011).
58. Kawakami, K., Amodio, D. M. & Hugenberg, K. in *Advances in Experimental Social Psychology* (ed. Olson, J. M.) Vol. 55, 1–80 (Elsevier, 2017).
59. Lin, C., Keles, U. & Adolphs, R. Four dimensions characterize attributions from faces using a representative set of English trait words. *Nat. Commun.* **12**, 5168 (2021).
60. Stolier, R. M. & Freeman, J. B. Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* **19**, 795–797 (2016).
61. Tamir, D. I., Thornton, M. A., Contreras, J. M. & Mitchell, J. P. Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. *Proc. Natl Acad. Sci. USA* **113**, 194–199 (2016).
62. Xie, S. Y., Flake, J. K. & Hehman, E. Perceiver and target characteristics contribute to impression formation differently across race and gender. *J. Pers. Soc. Psychol.* **117**, 364 (2019).
63. Amodio, D. M. in *Dual-process Theories of the Social Mind* (eds Sherman, J. W., Gawronski, B. & Trope, Y.) 560–577 (Guilford, 2014).
64. Hackel, L. M., Mende-Siedlecki, P., Loken, S. & Amodio, D. M. Context-dependent learning in social interaction: trait impressions support flexible social choices. *J. Pers. Soc. Psychol.* **123**, 655–675 (2022).
65. John, O. P., Hampson, S. E. & Goldberg, L. R. The basic level in personality-trait hierarchies: studies of trait use and accessibility in different contexts. *J. Pers. Soc. Psychol.* **60**, 348–361 (1991).

66. Kunda, Z., Sinclair, L. & Griffin, D. Equal ratings but separate meanings: stereotypes and the construal of traits. *J. Pers. Soc. Psychol.* **72**, 720–734 (1997).
67. Mann, T. C. & Ferguson, M. J. Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *J. Pers. Soc. Psychol.* **108**, 823–849 (2015).
68. Macrae, C. N. & MacLeod, M. D. On recollections lost: when practice makes imperfect. *J. Pers. Soc. Psychol.* **77**, 463–473 (1999).
69. Storm, B. C., Bjork, E. L. & Bjork, R. A. Social metacognitive judgments: the role of retrieval-induced forgetting in person memory and impressions. *J. Mem. Lang.* **52**, 535–550 (2005).
70. Okten, I. O. & Moskowitz, G. B. Easy to make, hard to revise: updating spontaneous trait inferences in the presence of trait-inconsistent information. *Soc. Cogn.* **38**, 571–625 (2020).
71. Tulving, E. Episodic memory: from mind to brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
72. Dickerson, B. C. & Eichenbaum, H. The episodic memory system: neurocircuitry and disorders. *Neuropsychopharmacology* **35**, 86–104 (2010).
73. Kadwe, P. P., Sklenar, A. M., Frankenstein, A. N., Levy, P. U. & Leshikar, E. D. The influence of memory on approach and avoidance decisions: investigating the role of episodic memory in social decision making. *Cognition* **225**, 105072 (2022).
74. Klein, S. B. et al. Evolution and episodic memory: an analysis and demonstration of a social function of episodic recollection. *Soc. Cogn.* **27**, 283–319 (2009).
75. Meiser, T. Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Soc. Cogn.* **21**, 121–156 (2003).
76. FeldmanHall, O., Montez, D. F., Phelps, E. A., Davachi, L. & Murty, V. P. Hippocampus guides adaptive learning during dynamic social interactions. *J. Neurosci.* **41**, 1340–1348 (2021).
77. Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A. & Davachi, L. Episodic memories predict adaptive value-based decision-making. *J. Exp. Psychol. Gen.* **145**, 548–558 (2016).
78. Hastie, R. & Park, B. The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychol. Rev.* **93**, 258–268 (1986).
79. Lupo, A. K. & Zárate, M. A. Guilty by association: time-dependent memory consolidation facilitates the generalization of negative — but not positive — person memories to group and self-judgments. *J. Exp. Soc. Psychol.* **83**, 78–87 (2019).
80. Brainerd, C. J. & Reyna, V. F. Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* **11**, 164–169 (2002).
81. Hackel, L. M. & Mende-Siedlecki, P. Two modes of social impressions and their effects on choice. *J. Exp. Psychol. Gen.* **152**, 3002–3020 (2023).
82. Enge, L. R., Lupo, A. K. & Zárate, M. A. Neurocognitive mechanisms of prejudice formation: the role of time-dependent memory consolidation. *Psychol. Sci.* **26**, 964–971 (2015).
83. Hupbach, A., Gomez, R. & Nadel, L. Episodic memory reconsolidation: updating or source confusion? *Memory* **17**, 502–510 (2009).
84. Wichert, S., Wolf, O. T. & Schwabe, L. Updating of episodic memories depends on the strength of new learning after memory reactivation. *Behav. Neurosci.* **127**, 331–338 (2013).
85. Loftus, E. F. & Hoffman, H. G. Misinformation and memory: the creation of new memories. *J. Exp. Psychol. Gen.* **118**, 100–104 (1989).
86. Schacter, D. L. et al. The future of memory: remembering, imagining, and the brain. *Neuron* **76**, 677–694 (2012).
87. Balcetis, E. Where the motivation resides and self-deception hides: how motivated cognition accomplishes self-deception. *Soc. Pers. Psychol. Compass* **2**, 361–381 (2008).
88. Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E. & Crockett, M. J. Motivated misremembering of selfish decisions. *Nat. Commun.* **11**, 2100 (2020).
89. Biernat, M. & Sesko, A. K. Communicating about others: motivations and consequences of race-based impressions. *J. Exp. Soc. Psychol.* **49**, 138–143 (2013).
90. Dodson, C. S., Darragh, J. & Williams, A. Stereotypes and retrieval-provoked illusory source recollections. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 460–477 (2008).
91. Taylor, S. E., Fiske, S. T., Etcoff, N. L. & Ruderman, A. J. Categorical and contextual bases of person memory and stereotyping. *J. Pers. Soc. Psychol.* **36**, 778–793 (1978).
92. Nunes, L. D., Garcia-Marques, L., Ferreira, M. B. & Ramos, T. Inferential costs of trait centrality in impression formation: organization in memory and misremembering. *Front. Psychol.* **8**, 1408 (2017).
93. Skinner, B. F. Operant behavior. *Am. Psychol.* **18**, 503–515 (1963).
94. Cox, J. & Witten, I. B. Striatal circuits for reward learning and decision-making. *Nat. Rev. Neurosci.* **20**, 482–494 (2019).
95. O'Doherty, J. et al. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
96. Liljeholm, M. & O'Doherty, J. P. Contributions of the striatum to learning, motivation, and performance: an associative account. *Trends Cogn. Sci.* **16**, 467–475 (2012).
97. Foerde, K. 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 (2018).
98. Robbins, T. W. & Costa, R. M. Habits. *Curr. Biol.* **27**, R1200–R1206 (2017).
99. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 1st edn (MIT Press, 1998).
100. Balleine, B. W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407–419 (1998).
101. Förster, J., Liberman, N. & Friedman, R. S. Seven principles of goal activation: a systematic approach to distinguishing goal priming from priming of non-goal constructs. *Pers. Soc. Psychol. Rev.* **11**, 211–233 (2007).
102. Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
103. Reber, P. J. & Squire, L. R. Parallel brain systems for learning with and without awareness. *Learn. Mem.* **1**, 217–229 (1994).
104. Knowlton, B. J., Squire, L. R. & Gluck, M. A. Probabilistic classification learning in amnesia. *Learn. Mem.* **1**, 106–120 (1994).
105. Graybiel, A. M. & Grafton, S. T. The striatum: where skills and habits meet. *Cold Spring Harb. Persp. Biol.* **7**, a021691 (2015).
106. Hackel, L. M., Doll, B. B. & Amodio, D. M. Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nat. Neurosci.* **18**, 1233–1235 (2015).
107. Ruff, C. C. & Fehr, E. The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* **15**, 549–562 (2014).
108. Lieberman, M. D. Intuition: a social cognitive neuroscience approach. *Psychol. Bull.* **126**, 109–137 (2000).
109. Hackel, L. M., Berg, J. J., Lindström, B. R. & Amodio, D. M. Model-based and model-free social cognition: investigating the role of habit in social attitude formation and choice. *Front. Psychol.* **10**, 2592 (2019).
110. Hackel, L. M., Mende-Siedlecki, P. & Amodio, D. M. Reinforcement learning in social interaction: the distinguishing role of trait inference. *J. Exp. Soc. Psychol.* **88**, 103948 (2020).
111. Hackel, L. M., Kogon, D., Amodio, D. M. & Wood, W. Group value learned through interactions with members: a reinforcement learning account. *J. Exp. Soc. Psychol.* **99**, 104267 (2022).
112. Traast, I. J., Schultner, D. T., Doosje, B. & Amodio, D. M. Race effects on impression formation in social interaction: an instrumental learning account. *J. Exp. Psychol. Gen.* **153**, 2985–3001 (2024).
113. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. *Class. Cond. Curr. Res. Theory* **2**, 64–69 (1972).
114. Rydell, R. J. & McConnell, A. R. Understanding implicit and explicit attitude change: a systems of reasoning analysis. *J. Pers. Soc. Psychol.* **91**, 995–1008 (2006).
115. Wood, W. & Neal, D. T. A new look at habits and the habit-goal interface. *Psychol. Rev.* **114**, 843–863 (2007).
116. Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* **7**, 464–476 (2006).
117. Wood, W. Habit in personality and social psychology. *Pers. Soc. Psychol. Rev.* **21**, 389–403 (2017).
118. Cho, H. J. & Hackel, L. M. Instrumental learning of social affiliation through outcome and intention. *J. Exp. Psychol. Gen.* **151**, 2204–2221 (2022).
119. Andreatta, M. & Pauli, P. Appetitive vs. aversive conditioning in humans. *Front. Behav. Neurosci.* **9**, 128 (2015).
120. Fendt, M. & Fanselow, M. S. The neuroanatomical and neurochemical basis of conditioned fear. *Neurosci. Biobehav. Rev.* **23**, 743–760 (1999).
121. Maren, S. Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.* **24**, 897–931 (2001).
122. Bechara, A. et al. Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* **269**, 1115–1118 (1995).
123. Morris, J. S., Öhman, A. & Dolan, R. J. Conscious and unconscious emotional learning in the human amygdala. *Nature* **393**, 467–470 (1998).
124. Roelofs, K. Freeze for action: neurobiological mechanisms in animal and human freezing. *Phil. Trans. R. Soc. B* **372**, 20160206 (2017).
125. Amodio, D. M., Harmon-Jones, E. & Devine, P. G. 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 (2003).
126. March, D. S., Gaertner, L. & Olson, M. A. On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychol. Inq.* **29**, 1–13 (2018).
127. Hütter, M. An integrative review of dual- and single-process accounts of evaluative conditioning. *Nat. Rev. Psychol.* **1**, 640–653 (2022).
128. Olson, M. A. & Fazio, R. H. Implicit attitude formation through classical conditioning. *Psychol. Sci.* **12**, 413–417 (2001).
129. Staats, A. W. & Staats, C. K. Attitudes established by classical conditioning. *J. Abnorm. Soc. Psychol.* **57**, 37–40 (1958).
130. Itkes, O. & Kron, A. Affective and semantic representations of valence: a conceptual framework. *Emot. Rev.* **11**, 283–293 (2019).
131. Rohr, M. & Wentura, D. How emotion relates to language and cognition, seen through the lens of evaluative priming paradigms. *Front. Psychol.* **13**, 911068 (2022).
132. Bouton, M. E. Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol. Bull.* **114**, 80–99 (1993).
133. Kindt, M., Soeter, M. & Vervliet, B. Beyond extinction: erasing human fear responses and preventing the return of fear. *Nat. Neurosci.* **12**, 256–258 (2009).
134. Monfils, M.-H., Cowansage, K. K., Klann, E. & LeDoux, J. E. Extinction–reconsolidation boundaries: key to persistent attenuation of fear memories. *Science* **324**, 951–955 (2009).
135. Schiller, D. et al. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* **463**, 49–53 (2010).
136. Elsey, J. W. B., Van Ast, V. A. & Kindt, M. Human memory reconsolidation: a guiding framework and critical review of the evidence. *Psychol. Bull.* **144**, 797–848 (2018).
137. Kindt, M. & Soeter, M. Reconsolidation in a human fear conditioning study: a test of extinction as updating mechanism. *Biol. Psychol.* **92**, 43–50 (2013).

138. Delgado, M. R., Olsson, A. & Phelps, E. A. Extending animal models of fear conditioning to humans. *Biol. Psychol.* **73**, 39–48 (2006).
139. Dunsmoor, J. E., Kubota, J. T., Li, J., Coelho, C. A. O. & Phelps, E. A. Racial stereotypes impair flexibility of emotional learning. *Soc. Cogn. Affect. Neurosci.* **11**, 1363–1373 (2016).
140. Olsson, A., Ebert, J. P., Banaji, M. R. & Phelps, E. A. The role of social groups in the persistence of learned fear. *Science* **309**, 785–787 (2005).
141. March, D. S., Gaertner, L. & Olson, M. A. Danger or dislike: distinguishing threat from negative valence as sources of automatic anti-Black bias. *J. Pers. Soc. Psychol.* **121**, 984 (2021).
142. Dovidio, J. F., Kawakami, K. & Gaertner, S. L. Implicit and explicit prejudice and interracial interaction. *J. Pers. Soc. Psychol.* **82**, 62–68 (2002).
143. Fazio, R. H., Jackson, J. R., Dunton, B. C. & Williams, C. J. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* **69**, 1013–1027 (1995).
144. Shelton, J. N. & Richeson, J. A. Interracial interactions: a relational approach. *Adv. Exp. Soc. Psychol.* **38**, 121–181 (2006).
145. Stephan, W. G. & Stephan, C. W. Intergroup anxiety. *J. Soc. Issues* **41**, 157–175 (1985).
146. Word, C. O., Zanna, M. P. & Cooper, J. The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *J. Exp. Soc. Psychol.* **10**, 109–120 (1974).
147. Moskowitz, G. B., Okten, I. O. & Schneid, E. in *The Handbook of Impression Formation* (eds. Balcetis, E. & Moskowitz, G. B.) ch. 18 (Routledge, 2022).
148. Phelps, E. A. Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* **14**, 198–202 (2004).
149. Poldrack, R. A. & Packard, M. G. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* **41**, 245–251 (2003).
150. Greenwald, A. G. & Banaji, M. R. The implicit revolution: reconceiving the relation between conscious and unconscious. *Am. Psychol.* **72**, 861–871 (2017).
151. Gawronski, B., Ledgerwood, A. & Eastwick, P. W. Implicit bias ≠ bias on implicit measures. *Psychol. Inq.* **33**, 139–155 (2022).
152. Corneille, O. & Hütter, M. Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Pers. Soc. Psychol. Rev.* **24**, 212–232 (2020).
153. Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C. & Kardes, F. R. On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* **50**, 229–238 (1986).
154. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
155. Payne, B. K., Cheng, C. M., Govorun, O. & Stewart, B. D. An inkblot for attitudes: affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* **89**, 277–293 (2005).
156. Amodio, D. M. & Devine, P. G. Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *J. Pers. Soc. Psychol.* **91**, 652–661 (2006).
157. Blaison, C., Imhoff, R., Hühnel, I., Hess, U. & Banse, R. The affect misattribution procedure: hot or not? *Emotion* **12**, 403–412 (2012).
158. Argaman, Y., Heimer, O., Bar-Anan, Y. & Kron, A. Does the Brief Implicit Association Test measure semantic or affective valence representations? *Emotion* https://doi.org/10.1037/emo0001480 (2024).
159. Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A. & Kron, A. Dissociating affective and semantic valence. *J. Exp. Psychol. Gen.* **146**, 924–942 (2017).
160. Wentura, D. & Degner, J. Automatic evaluation isn't that crude! Moderation of masked affective priming by type of valence. *Cogn. Emot.* **24**, 609–628 (2010).
161. Wittenbrink, B., Judd, C. M. & Park, B. Spontaneous prejudice in context: variability in automatically activated attitudes. *J. Pers. Soc. Psychol.* **81**, 815–827 (2001).
162. Kret, M. E. Emotional expressions beyond facial muscle actions. a call for studying autonomic signals and their impact on social perception. *Front. Psychol.* **6**, 711 (2015).
163. Clore, G. L. & Gormly, J. B. Knowing, feeling, and liking: a psychophysiological study of attraction. *J. Res. Pers.* **8**, 218–230 (1974).
164. Schultner, D. T. et al. Transmission of societal stereotypes to individual-level prejudice through instrumental learning. *Proc. Natl Acad. Sci. USA* **121**, e2414518121 (2024).
165. Ferguson, M. J. & Bargh, J. A. Liking is for doing: the effects of goal pursuit on automatic evaluation. *J. Pers. Soc. Psychol.* **87**, 557–572 (2004).
166. Strack, F. & Deutsch, R. Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* **8**, 220–247 (2004).
167. Allport, G. W. *The Nature of Prejudice* xviii, 537 (Addison-Wesley, 1954).
168. Amodio, D. M. & Ratner, K. G. A memory systems model of implicit social cognition. *Curr. Dir. Psychol. Sci.* **20**, 143–148 (2011).
169. Dovidio, J. F., Esses, V. M., Beach, K. R. & Gaertner, S. L. in *From Prejudice to Intergroup Emotions* (eds. Mackie, D. M. & Smith, E. R.) ch. 9 (Psychology Press, 2002).
170. Amodio, D. M. & Hamilton, H. K. Intergroup anxiety effects on implicit racial evaluation and stereotyping. *Emotion* **12**, 1273–1280 (2012).
171. Bijlstra, G., Holland, R. W. & Wigboldus, D. H. J. The social face of emotion recognition: evaluations versus stereotypes. *J. Exp. Soc. Psychol.* **46**, 657–663 (2010).
172. Wittenbrink, B., Judd, C. M. & Park, B. Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *J. Pers. Soc. Psychol.* **72**, 262–274 (1997).
173. Olcaysoy Okten, I., Schneid, E. D. & Moskowitz, G. B. On the updating of spontaneous impressions. *J. Pers. Soc. Psychol.* **117**, 1–25 (2019).
174. Kurdi, B., Mann, T. C., Charlesworth, T. E. S. & Banaji, M. R. The relationship between implicit intergroup attitudes and beliefs. *Proc. Natl Acad. Sci. USA* **116**, 5862–5871 (2019).
175. Phills, C. E., Hahn, A. & Gawronski, B. The bidirectional causal relation between implicit stereotypes and implicit prejudice. *Pers. Soc. Psychol. Bull.* **46**, 1318–1330 (2020).
176. Roelofs, K., Hagenaars, M. A. & Stins, J. Facing freeze: social threat induces bodily freeze in humans. *Psychol. Sci.* **21**, 1575–1581 (2010).
177. Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B. & Howard, A. On the nature of prejudice: automatic and controlled processes. *J. Exp. Soc. Psychol.* **33**, 510–540 (1997).
178. Hupbach, A., Olcaysoy Okten, I. & Horn, P. Directed forgetting in the social domain: forgetting behaviors but not inferred traits. *J. Appl. Res. Mem. Cogn.* **11**, 522–533 (2022).
179. Foerde, K., Knowlton, B. J. & Poldrack, R. A. Modulation of competing memory systems by distraction. *Proc. Natl Acad. Sci. USA* **103**, 11778–11783 (2006).
180. Foerde, K. & Shohamy, D. Feedback timing modulates brain systems for learning in humans. *J. Neurosci.* **31**, 13157–13167 (2011).
181. Foerde, K., Race, E., Verfaellie, M. & Shohamy, D. A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. *J. Neurosci.* **33**, 5698–5704 (2013).
182. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
183. Nicholas, J., Daw, N. D. & Shohamy, D. Uncertainty alters the balance between incremental learning and episodic memory. *eLife* **11**, e81679 (2022).
184. Doll, B. B., Jacobs, W. J., Sanfey, A. G. & Frank, M. J. Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Res.* **1299**, 74–94 (2009).
185. Lindström, B., Golkar, A., Jangard, S., Tobler, P. N. & Olsson, A. Social threat learning transfers to decision making in humans. *Proc. Natl Acad. Sci. USA* **116**, 4732–4737 (2019).
186. Kensinger, E. A. Remembering the details: effects of emotion. *Emot. Rev.* **1**, 99–113 (2009).
187. LaBar, K. S. & Phelps, E. A. Arousal-mediated memory consolidation: role of the medial temporal lobe in humans. *Psychol. Sci.* **9**, 490–493 (1998).
188. McGaugh, J. L. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* **27**, 1–28 (2004).
189. Skowronski, J. J. & Carlston, D. E. Negativity and extremity biases in impression formation: a review of explanations. *Psychol. Bull.* **105**, 131–142 (1989).
190. Fazio, R. H., Eiser, J. R. & Shook, N. J. Attitude formation through exploration: valence asymmetries. *J. Pers. Soc. Psychol.* **87**, 293–311 (2004).
191. Denrell, J. Why most people disapprove of me: experience sampling in impression formation. *Psychol. Rev.* **112**, 951–978 (2005).
192. Allidina, S. & Cunningham, W. A. Avoidance begets avoidance: a computational account of negative stereotype persistence. *J. Exp. Psychol. Gen.* **150**, 2078–2099 (2021).
193. Bai, X., Fiske, S. T. & Griffiths, T. L. Globally inaccurate stereotypes can result from locally adaptive exploration. *Psychol. Sci.* **33**, 671–684 (2022).
194. Cone, J. & Ferguson, M. J. He did what? The role of diagnosticity in revising implicit evaluations. *J. Pers. Soc. Psychol.* **108**, 37–57 (2015).
195. Olcaysoy Okten, I. & Moskowitz, G. B. Understanding impression updating through behavior predictions. *Soc. Cogn.* **42**, 198–232 (2024).
196. Hackel, L. M. & Amodio, D. M. Computational neuroscience approaches to social cognition. *Curr. Opin. Psychol.* **24**, 92–97 (2018).
197. Lockwood, P. L. & Klein-Flügge, M. C. Computational modelling of social cognition and behaviour — a reinforcement learning primer. *Soc. Cogn. Affect. Neurosci.* **16**, 761–771 (2021).
198. Cone, J., Flaharty, K. & Ferguson, M. J. Believability of evidence matters for correcting social impressions. *Proc. Natl Acad. Sci. USA* **116**, 9802–9807 (2019).
199. Cone, J. & Calanchini, J. A process dissociation model of implicit rapid revision in response to diagnostic revelations. *Pers. Soc. Psychol. Bull.* **47**, 201–215 (2021).
200. Melnikoff, D. E. & Bargh, J. A. The mythical number two. *Trends Cogn. Sci.* **22**, 280–293 (2018).
201. Carew, T. J., Hawkins, R. D. & Kandel, E. R. Differential classical conditioning of a defensive withdrawal reflex in aplysia californica. *Science* **219**, 397–400 (1983).
202. De Houwer, J. Using the implicit association test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learn. Motiv.* **37**, 176–187 (2006).
203. Gregg, A. P., Seibt, B. & Banaji, M. R. Easier done than undone: asymmetry in the malleability of implicit preferences. *J. Pers. Soc. Psychol.* **90**, 1–20 (2006).
204. Hahn, A., Judd, C. M., Hirsh, H. K. & Blair, I. V. Awareness of implicit attitudes. *J. Exp. Psychol. Gen.* **143**, 1369–1392 (2014).
205. Morris, A. & Kurdi, B. Awareness of implicit attitudes: large-scale investigations of mechanism and scope. *J. Exp. Psychol. Gen.* **152**, 3311–3343 (2023).
206. Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* **306**, 1940–1943 (2004).
207. Bechara, A., Damasio, H., Damasio, A. R. & Lee, G. P. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* **19**, 5473–5481 (1999).
208. Solarz, A. K. Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *J. Exp. Psychol.* **59**, 239–245 (1960).
209. Kurdi, B. & Dunham, Y. Propositional accounts of implicit evaluation: taking stock and looking ahead. *Soc. Cogn.* **38**, s42–s67 (2020).
210. Schultner, D. T., Lindström, B. R., Cikara, M. & Amodio, D. M. Transmission of social bias through observational learning. *Sci. Adv.* **10**, eadk2030 (2024).
211. Amodio, D. M. & Cikara, M. The social neuroscience of prejudice. *Annu. Rev. Psychol.* **72**, 439–469 (2021).
212. Payne, B. K., Burkley, M. A. & Stokes, M. B. Why do implicit and explicit attitude tests diverge? The role of structural fit. *J. Pers. Soc. Psychol.* **94**, 16–31 (2008).

213. Phelps, E. A. et al. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* **12**, 729–738 (2000).
214. Rösler, I. K. & Amodio, D. M. Neural basis of prejudice and prejudice reduction. *Biol. Psychiat. Cogn. Neurosci. Neuroimag.* **7**, 1200–1208 (2022).
215. Macrae, C. N. & Bodenhausen, G. V. Social cognition: thinking categorically about others. *Annu. Rev. Psychol.* **51**, 93–120 (2000).
216. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
217. Botvinick, M. M., Cohen, J. D. & Carter, C. S. Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* **8**, 539–546 (2004).
218. Aron, A. R., Robbins, T. W. & Poldrack, R. A. Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* **8**, 170–177 (2004).
219. Badre, D. Cognitive control. *Annu. Rev. Psychol.* **76**, 167–195 (2025).
220. Gollwitzer, P. M. Implementation intentions: strong effects of simple plans. *Am. Psychol.* **54**, 493–503 (1999).
221. Amodio, D. M. & Swencionis, J. K. Proactive control of implicit bias: a theoretical model and implications for behavior change. *J. Pers. Soc. Psychol.* **115**, 255 (2018).
222. Monteith, M. J., Ashburn-Nardo, L., Voils, C. I. & Czopp, A. M. Putting the brakes on prejudice: on the development and operation of cues for control. *J. Pers. Soc. Psychol.* **83**, 1029–1050 (2002).
223. Monteith, M. J., Sherman, J. W. & Devine, P. G. Suppression as a stereotype control strategy. *Pers. Soc. Psychol. Rev.* **2**, 63–82 (1998).
224. Wegner, D. M. Ironic processes of mental control. *Psychol. Rev.* **101**, 34–52 (1994).

Acknowledgements
This work was supported by a Vici grant (016.185.058) from the Netherlands Organisation for Scientific Research. The author thanks L. Hackel, K. Foerde, M. Kindt and M. Vollberg for their valuable comments on previous versions of this article.

Competing interests
The author declares no competing interests.

Additional information
Peer review information *Nature Reviews Psychology* thanks Irmak Okten, who co-reviewed with Ayanna Brewton, and Gordon Moskowitz for their contribution to the peer review of this work.