PREPRINT

Learning mechanisms underlying impression formation and updating

David M. Amodio, University of Amsterdam

Correspondence:
David M. Amodio
Dept. of Psychology
Nieuwe Achtergracht 129
1001 NK Amsterdam
Email: d.m.amodio@uva.nl

Conditionally accepted at *Nature Reviews Psychology*; please note that this version does not include final edits or final figures.

Abstract

Impression formation is the process of learning about people—how we infer a person's character traits, goals, and preferences while forming our own attitudes toward them. Emerging research shows that impressions are formed through a variety of mechanisms—a multimodal process rooted in different underlying systems of learning and memory. In this review, 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 our understanding of how impressions are formed, changed, and expressed in behavior, relative to prior accounts based only on semantic memory models, while illuminating longstanding debates on the nature of implicit social cognition and how social information is represented in the mind.

I. Introduction

> "This remarkable capacity we possess to understand something of the character of another person … is a precondition of social life."
>
> (Asch, 1946, p. 258)

As humans, we depend on other to survive and thrive, and our ability to assess people—to infer their traits and motives and discern friend from foe—is a fundamental capacity of the human mind (Asch, 1946). 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.

Social impressions have long been considered multifaceted (Carlston, 1994; Jones & Davis, 1965; Malle & Holbrook, 2012). They involve conceptual knowledge of a target person's attributes, such as their trait characteristics (Asch, 1946; Trope, 1986; Gilbert et al., 1988; Winter & Uleman, 1994), goals and intentions (Hassin et al., 2005; Heider, 1944; Read et al., 1990; Moskowitz & Olcaysoy Okten, 2018), mental states (Ames, 2004; Kruse & Degner, 2021), and life circumstances (e.g., wealth, geography, group memberships; Brewer, 1988; Fiske & Neuberg, 1990; Kunda & Thagard, 1996). Impressions also involve a perceiver's own attitude toward a target (Anderson, 1965; Cone et al., 2017; Schneid et al., 2015), which may include their evaluative beliefs and associations, affective responses, and behavioral dispositions (e.g., to approach or avoid) (Breckler, 1984). These varied aspects of an impression reflect the multiple ways in which humans learn, through semantic, episodic, instrumental, and Pavlovian memory systems, and how these learning processes function together to guide social judgments, decisions, and actions (Amodio, 2019). These learning mechanisms further guide how impressions are changed (i.e., updated) in response to new information and experiences. Although classic accounts of impression formation emphasize conceptual inferences rooted in in semantic memory (Anderson, 1965; Asch, 1946; Cantor & Mischel, 1977; Hastie et al., 1980), it is now clear that multiple learning and memory mechanisms contribute to how we think about and act toward people.

In this review, 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 processes involved in social cognition—their content, modes of acquisition and change, functions, and expressions—and discuss their implications for impression formation and updating. I then discuss how a consideration of these learning mechanisms, and their interactions, illuminate longstanding theoretical questions regarding the nature of implicit attitudes and process models of social cognition.

II. Learning and memory systems

When we meet someone, we experience them simultaneously in multiple ways: we encode the details of the event (e.g., their appearance and nonverbals, other people involved, the context), infer their trait attributes and goals, react emotionally to their feedback, and track how they respond—positively or negatively—to things we do and say (Figure 1). Each form of learning—the episodic details, the traits we infer, the responses we track, the emotion we experience—contributes to our emergent impression of that person.

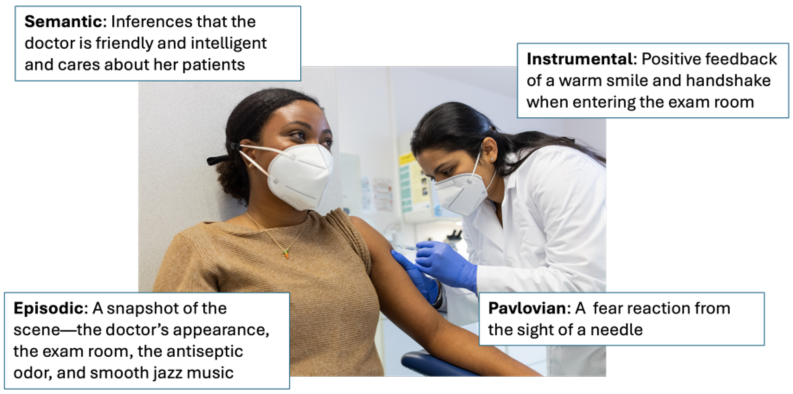


Figure 1. As a perceiver forms an impression, they simultaneously encode information through mulitple memory systems. For example, when meeting a doctor for a vaccination, we may infer her traits as intelligent and caring (semantic), form reward associations from her positive feedback (instrumental), form a fear association when spotting the needle (Pavlovian), all while encoding the multimodal details of the situation (episodic).

The idea that human thought and behavior are rooted in mechanisms of learning and memory is foundational in psychological science (Collins & Loftus, 1974; Hull, 1943; Pavlov, 1927; Scoville and Milner, 1957; Shiffrin & Schneider, 1977; Thorndike, 1932; Tolman, 1948), and it inspired the emergence of social cognition—a field originally known as "person memory" (Hastie et al., 1980). 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 (Gabrieli, 1998; McDonald & White, 1993; Squire & Zola, 1996; Tulving, 1985). Although distinct memory system functions are revealed most dramatically in studies of selective brain damage (Bechara et al., 1995; Knowlton et al., 1996; Scoville & Milner, 1957), they typically operate in concert in the healthy mind (Amodio, 2019; Henke, 2010; Squire, 2004).

In this section, I describe major learning and memory systems that are most relevant to impression formation. I highlight the specific kinds of information they encode, 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 judgment and behavior.

Semantic memory. Semantic memory refers to the learning, representation, and retrieval of general knowledge—*the sky is blue*, *1 + 1 = 2*, and *my friend Sally is friendly, smart, and athletic*. Early theories of person perception and social cognition were inspired by models of semantic memory (Hastie et al., 1980; Uleman & Kressel, 2013), and contemporary models of impression formation continue to assume a basis in semantic processes (Amodio, 2019; Moskowitz, 2024).

Semantic memory is declarative, such that it is explicitly reportable, and propositional, in that it meaningfully links abstract linguistic concepts (Kumar, 2021). Semantic memory is primarily represented in the anterior temporal lobe (Binney & Ramsey, 2020; Olson et al., 2013) and activated during social judgments in the medial frontal cortex (Contreras et al., 2012; Gilbert et al., 2012). Although typically expressed via verbal self-report, semantic associations can be expressed on indirect measures involving conceptual categorization (e.g., semantic priming). That is, while a perceiver is aware of semantic knowledge, this knowledge may be expressed indirectly (i.e., implicitly) and thus 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 (Anderson, 1965; Asch, 1946; Moskowitz & Olcaysoy Okten, 2018; Read et al., 1990), encoded as cognitive concepts organized in a semantic network (Collins & Loftus, 1975; Wyer, 1980). Semantic impressions may be formed through direct verbal descriptions of a person (Asch, 1946) or inferred from a person's behavior (Carlston & Skowronki, 1994; Jones & Davis, 1965; Srull & Wyer, 1979; Winter & Uleman, 1984).

When semantic knowledge is activated, such as when encountering a target individual, this information becomes accessible and can influence person judgments (Higgins et al., 1977; Bargh & Pietromonaco, 1982). This semantic form of person knowledge underpins major theories of implicit social cognition (Fazio, 1990; Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995; Kunda & Thagard, 1996; Smith & DeCoster, 2000) and intergroup bias (Devine, 1989; Hamilton & Sherman, 1994; Kawakami et al., 2017; Sherman, 1996), as well as more recent models of intersectional and multidimensional impression formation (Chen, 2014; Freeman & Ambady, 2011; Lin, Keles, & Adolphs, 2021; Stolier & Freeman, 2016; Tamir et al., 2016).

It's unsurprising that theories of impression formation are dominated by semantic models. Being declarative, semantic information is most salient in the mind of a social perceiver (Amodio, 2014). Moreover, semantic knowledge is highly functional in a complex social millieu, as it affords precision, nuance, and flexibility; drawn from a rich descriptive lexicon, complex semantic impressions can describe a person from multiple angles and across contexts (Hackel et al., 2022a; John, Hampson, & Goldberg, 1991).

Updating of semantic knowledge occur not through change per se, but through elaboration based on new learning (Kunda, Sinclair, & Griffin, 1997). In the context of impressions, one may learn new complementary or contradictory trait information. Existing knowledge may also be reinterpreted in light of new information (Mann & Ferguson, 2015), revised during retrieval (Storm, Bjork, & Bjork, 2005) or, when no longer relevant, forgotten (Dunn & Spellman, 2003; Macrae & MacLeod, 1999). Although old trait information is typically retained alongside new knowledge, a perceiver can select relevant new information when forming explicit judgments or summary evaluations (Olcaysoy Okten et al., 2019; Olcaysoy Okten & Moskowitz, 2020). Thus, while your impression of Bob, the junk hoarding neighbor, improves when you learn he recycles toys for sick children, your knowledge of him as a hoarder remains.

Episodic memory. Episodic memory encodes multimodal snapshots of our discrete experiences, from the extraordinary—the moment in the delivery room when you first set eyes on your newborn child—to the mundane, like yesterday's lunch transaction at the local deli (Tulving, 2002). Early evidence that episodic memory functions as an independent system came from studies of brain lesion patients. In the famous case of patient H.M., the removal of his 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 piano—capacities that rely on semantic and instrumental memory (Scoville & Milner, 1957). Since then, studies of the medial temporal lobe in brain lesion patients and in healthy individuals, using neuroimaging, have further established episodic memory as a separable memory system (Dickerson & Eichenbaum, 2010; Baddely, 2001).

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, his fiancé's surprised look, the sparkle of a diamond, and the cheering crowd. Episodic memories provide a basis for trait inference with specific examples of a person's behavior (Kadwe et al., 2022; Klein et al. 2009; Meiser, 2003). For example, episodic recall of how much a person shared in a prior interaction relies on the hippocampus (FeldmanHall et al., 2021) and informs a perceiver's choice of whether to engage with that person again (Murty et al., 2016).

Episodic memory can also provide a basis for semantic inference (Hastie & Park, 1986). A perceiver can infer trait characteristics from episodes of a person's behavior, deliberatively through attribution (Jones & Davies, 1965) or automatically through spontaneous trait inferences (Winter & Uleman, 1984). Similarly, a discrete episodic memory (e.g., vividly recalling 100 people at an event) can give rise to "gist" memories ("there was a big crowd"), with both simultaneously encoded (Brainerd & Reyna, 2002). When making social decisions, episodes guide specific judgments whereas gist guides more general, flexible judgments (Hackel & Mende-Siedlecki, 2023).

The updating of episodic memory is not incremental, but involves reconsolidation and the integration of new information (Hupbach et al., 2009; Wichert et al., 2013). Depending on a perceiver's goals or the salience of the episodes, the newer episodes may weigh more heavily in a perceiver's impression or decision (Bornstein et al., 2017; Hackel & Mende-Siedlecki, 2023; Kensinger & Corkin, 2004). Episodic memories can also be distorted through simulation, misremembering, retrospective reconstrual, reconsolidation, or imagined events (Anderson & Hanslmayr, 2014; Enge et al., 2015; Loftus & Hoffman, 1989; Hupbach et al., 2009; Schacter et al., 2012), often in ways that support a stereotype, schema, or self interest (Balcetis, 2008; Biernat & Sesko, 2013; Carlson et al., 2020; Dodson et al., 2008; Nunes et al., 2017; Taylor et al., 1978). In this way, newer or distorted episodes can contribute to a change in one's overall person impression.

Instrumental learning. Instrumental learning (also known as *operant conditioning* or *procedural memory*) is an action-based form of learning in which behaviors are associated with outcomes through reinforcement (Skinner, 1963; Thorndike, 1932). In contrast to semantic and episodic memories, instrumental learning is encoded in terms of reward value via dopaminergic activity in the striatum, and expressed directly in behavior (O'Doherty et al., 2004; Liljeholm & O'Doherty, 2012). Instrumental learning encompasses both goal-directed learning, which supports intentional, reward-driven behavior, and habits, which support automatically-cued responses (Foerde, 2018; Robbins & Costa, 2017).

Goal-directed instrumental learning. In goal-directed instrumental learning, one learns the reward value of an action—such as approaching an object or person—through choice and feedback. Following rules of reward reinforcement learning, choices that result in positive feedback are repeated and those resulting in negative feedback are not (Sutton & Barto, 1998). Instrumental learning occurs incrementally, such that reward associations change slowly through repeated experiences with action and feedback, and it can encode probabilistic reward contingencies (Balleine & Dickinson, 1998). Given its capacity to encode and express preferences through action, instrumental learning has been theorized to underlie the behavioral (or *conative*) component of attitudes (Amodio, 2019; Breckler, 1984) and the priming of goal-directed behavior (Forster, Liberman, & Friedman, 2007).

Instrumental learning is further distinguished by its *nondeclarative*, or implicit, operation, such that its associations may be formed and expressed without deliberation or awareness (Knowlton et al., 1996; Reber & Squire, 1994). For example, on probabilistic reinforcement tasks that involve incremental learning and thus require the ability to track accumulated feedback across many trials, amnesiac patients, who lack hippocampal function but retain normal striatal function, learn to make correct behavioral choices but are unaware of what they learned (Knowlton et al., 1994). In healthy individuals, nondeclarative instrumental learning is often expressed as a skill—a well-practiced, goal-directed action sequence that proceeds with little thought, such as playing piano, driving standard transmission, or swinging a golf club (Graybiel & Grafton, 2015).

In the context of impression formation, instrumental learning governs how we learn about others through direct social interaction—that is, through the exchange of action and feedback with another person (Amodio, 2019; Hackel et al., 2015; Ruff & Fehr, 2014). Research that has combined behavioral experiments with computational modeling and fMRI shows that through direct social interaction, perceivers encode the reward value of choosing a partner in addition to inferring the partner's trait characteristics, and that these separate representations, encoded in different patterns of neural activity, have joint effects on social decisions (Hackel et al., 2015). Instrumental associations have been likened to a gut feeling or intuition (Lieberman, 2000) and, as a component of person impressions, they function implicitly to guide social choices independently of explicit traits or attitudes (Hackel et al., 2019, 2020, 2022a, 2022b; Cho & Hackel, 2022; Traast et al., 2024).

As noted above, instrumental associations are updated incrementally in response to prediction errors to maintain a running representation of a reward-based preference (i.e., expected value) (Rescorla & Wagner, 1972; Sutton & Barto, 1998). 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 and learning rate (i.e., the weighting of new information), resulting in a revised expected value—a form of incremental preference updating that closely resembles the kind of evaluative change examined in studies of impression updating.

A benefit of this instrumental learning approach is that the parameters representing these updating mechanisms can be quantified, along with other psychological factors of interest, in formalized computational model of how reward value is formed, updated, and expressed (Sutton & Barto, 1998). These models can then be tested by assessing the fit of human behavioral data (e.g., from an experimental task) to model-simulated data. This powerful approach to theory testing is increasingly used in social cognition research (Cushman et al., 2023; FeldmanHall & Nassar, 2021; Hackel & Amodio, 2018).

Habits. Frequently enacted behaviors, whether goal-directed or not, can transform into a *habit*—a behavior that is automatically triggered in response to an associated cue despite being contradictory or irrelevant to one's goals (Wood & Neal, 2007). Whereas goal-directed instrumental learning is associated with reward processing in the ventral striatum, habits are associated with dorsal striatum activity (Foerde, 2018; Robbins & Costa, 2017; Yin & Knowlton, 2006).

In social contexts, habits are expressed when a person's presence, actual or symbolic, activates an automatic behavioral response (Amodio, 2019; Hackel et al., 2019; Wood, 2017). Such habits can be adaptive: they can enhance the fluency of social interactions, requiring few cognitive resources, and an impression based in habit may be more resistant to inconsistencies in a partner's behavior. However, habits may become maladaptive when a partner or relationship changes: you may offer a beer to a friend who recently quit drinking or mindlessly text an old flame post-breakup. In either case, a habit's indifference 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 (Cho & Hackel, 2022; Hackel et al., 2015, 2019, 2022b).

Pavlovian learning (classical conditioning)

Édouard Claparède, the Swiss neurologist, famously described a patient with severe amnesia who greeted him each day as if they had never met. As the story goes, one day, in 1911, he 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 in 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 a learned association between a neutral stimulus and autonomically-arousing threat or reward (Rescorla, 1988). Although Pavlovian learning can be aversive or appetitive, most research in humans and animals has focused on aversive (i.e., fear) conditioning. Pavlovian learning is differentiated from other memory systems by its unique characteristics and substrate in amygdala circuitry (Fendt & Fanselow, 1999; Maren, 2001). Pavlovian aversive conditioning can occur nonconsciously (Bechara et al., 1995; Öhman, 1998) and is expressed primarily as behavioral freezing, autonomic arousal, and heightened attentional vigilance (Roelofs, 2017).

It is notable that the term "classical conditioning" is sometimes invoked to describe evaluative conditioning in attitudes and impression research (Arenson et al., 1982; Olson & Fazio, 2004; Staats & Staats, 1958); however, evaluative conditioning procedures typically involve the pairing of two conceptual stimuli (e.g., words) and do not typically elicit the physiological response associated with an amygdala-mediated Pavlovian learning mechanism. That is, they are Pavlovian in procedure but not mechanism, likely involving semantic learning rather than Pavlovian learning.

Pavlovian-conditioned associations are not directly updated; rather, new associations may be formed alongside existing associations (Bouton, 1993). 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 reestablished. Recent research suggests that it may be possible to change Pavlovian associations through reactivation and reconsolidation (Kindt et al., 2009; Monfils et al., 2009; Schiller et al., 2010), but it remains unclear whether this intervention changes the underlying association or only its expression in behavior (Elsey et al., 2018; Kindt & Soeter, 2013).

How does Pavlovian conditioning contribute to impression formation? Although aversive conditioning is robust in humans (Delgado et al., 2006) and has been proposed as a component of intergroup bias (Amodio et al., 2003; Dunsmoor et al., 2016; Olsson et al., 2005; March et al., 2018), 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. In studies of intergroup interaction, expressions of a Pavlovian form of prejudice appears evident in perceivers' social distance, stilted speech and action, interaction anxiety, and fear-related affect—much like Claparède's famous patient (Dovidio et al., 2002; Fazio et al., 1995; Shelton & Richeson, 2006; Stephan & Stephan, 1985; Word et al., 1974; Amodio & Hamilton, 2012; Cottrell & Neuberg, 2005). Although more research is needed to determine the role of Pavlovian learning in social impressions, these findings suggest it supports affective and threat-related behavioral responses to persons and groups.

Section summary: A memory systems model of impression formation. A memory systems analysis clarifies that we learn about and represent persons through multiple learning mechanisms: semantic, episodic, instrumental, and Pavlovian. As illustrated in Figure 2, these memory systems are separable, with unique operating characteristics and distinct neural substrates, and they function to produce specific kinds of social behavior. The multilevel person representation they create constitutes an impression—a collection of knowledge, beliefs, preferences, and opinions, as well as affective reactions and approach tendencies that produce our holistic view of a person. Although impressions have long been considered multifaceted (Asch, 1946; Carlston, 1994; Moskowitz et al., 2023), this analysis specifies the mechanisms supporting these facets and their unique roles in social behavior.

Independent and interactive effects of memory systems

Despite their unique features, learning and memory systems typically function in concert: during impression formation, we can simultaneously encode episodic information about the event, infer semantic knowledge about the person's traits and characteristics, develop a behavioral disposition through instrumental feedback, and form affective associations through Pavlovian learning (Amodio, 2019). Learning and memory systems also interact, whereby one memory system shapes or competes with another (Phelps, 2004; Poldrack & Packard, 2003). A consideration of these independent and interactive effects, and their influence on decisions, is essential for understanding how impressions are expressed in judgment and behavior. In this section, I describe examples of joint and interactive memory system effects and their implications for person impressions.

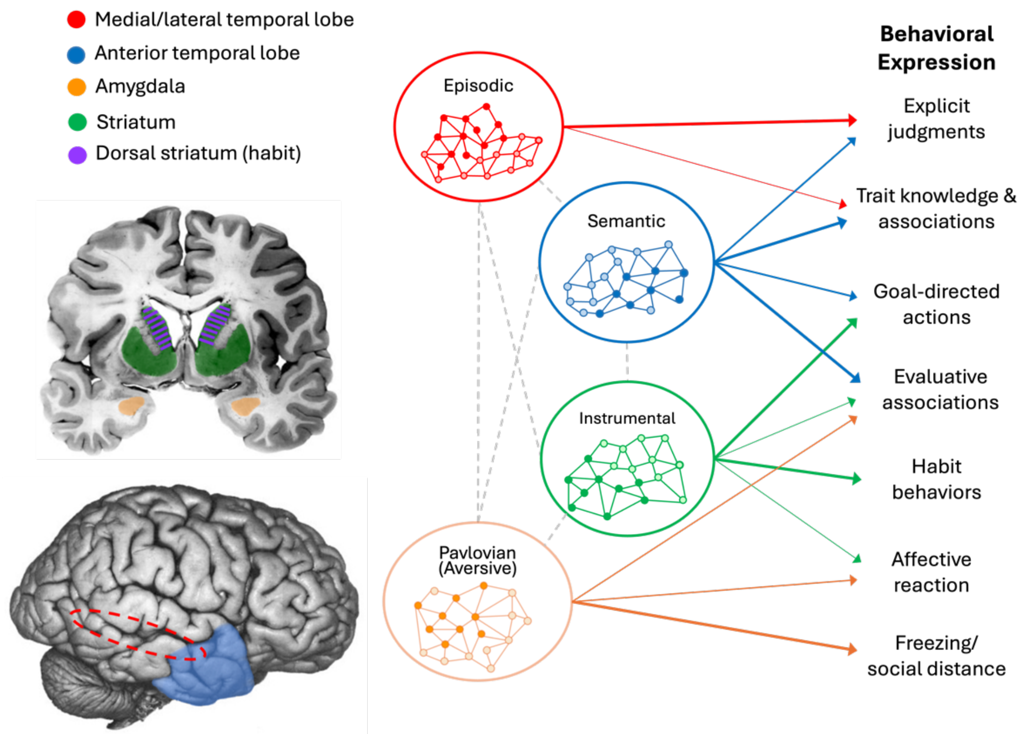


Figure 2. A memory systems model of person impression depicting episodic, semantic, instrumental (including habit), and Pavlovian aversive memory systems, their interconnectivity, neural correlates, and examples of their expressions in social behavior. A person impression may comprise one or more of these memory systems, and each may have varying degrees of influence on behavioral expressions (indicated by thickness of the arrows).

Independent effects of memory systems in person impressions. Independent effects refer to cases where two or more memory systems have simultaneous yet unique effects on judgment or behavior. I describe examples of such effects here.

*Multiple forms of implicit evaluation***.** Implicit evaluation refers to the indirect (i.e., nondeclarative) expression of positive or negative evaluation toward a person or object (Greenwald & Banaji, 2017), often assessed using tasks such as evaluative priming measures (e.g., Fazio et al., 1986), the Implicit Association Test (IAT) (Greenwald et al., 1998), or the Affect Misattribution Task (AMP) (Payne et al., 2005). Although a central to impression formation research, the construct of implicit evaluations—that is, how they are formed, represented in the mind, and expressed in behavior, and whether they function automatically or unconsciously—has been difficult to explain (Gawronski et al., 2022; Cornielle & Hutter, 2020).

From a learning perspective, implicit evaluation reflects the operation of one or more different underlying memory systems. For example, it could reflect an instrumental reward or Pavlovian threat association, both of which operate nondeclaratively and are expressed implicitly, or semantic knowledge which, although declarative and thus 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, expression, and potential for change.

Nearly all existing studies of implicit evaluation concern semantic memory. This is due to their reliance on tasks that primarily assess semantic associations between concepts and categories, such as evaluative priming tasks, the IAT, and AMP. Although some early models of implicit evaluation proposed a basis in affect (Amodio & Devine, 2006; Fazio et al., 1986; Gawronski & Bodenhausen, 2006), evidence for these accounts relied on data from semantic categorization tasks that, in subsequent work, have been shown to reflect semantic associations and not affective associations (Blaison et al, 2012). Thus, conventional implicit evaluation tasks, which rely on semantic categorization, are now understood to be primarily sensitive to semantic and not affective associations (De Houwer et al., 1998; Itkes et al., 2017; Klauer, 1997; Rohr & Wentura, 2022; Spruyt et al., 2004; Wentura and Degner, 2010; Wittenbrink et al., 2001). As such, they can further be understood as reflecting knowledge that is declarative but, when assessed with an implicit task, observed indirectly.

Affect-based implicit evaluation has been proposed to correspond to a Pavlovian association (Amodio et al., 2003; Amodio, 2019), which may be assessed by physiological measures of skin conductance or the startle eyeblink response (Kret, 2015). In early research on impression formation, heightened skin conductance response, an autonomic arousal indicator of either positive or negative affect depending on the elicitor, predicted greater attraction toward agreeable partners (Clore & Gormly, 1974). In the intergroup domain, my colleagues and I used a startle eyeblink method—an index of amygdala activity associated with the Pavlovian threat response—to assess White American participants' implicit affective responses to Black, White, and Asian faces (Amodio et al., 2003). We found that the startle response was amplified 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, which functions nondeclaratively and is expressed in physiological arousal and defensive behaviors, distinct from implicit evaluations based in semantic memory.

A third form of implicit evaluation is represented by Instrumental reward associations. Recent research shows that individuals form preferences for people through instrumental learning, using probabilistic reinforcement learning tasks in which participants choose to interact with individuals and receive either reward or nonreward feedback (Hackel et al., 2015, 2020, 2022a,b; Traast et al., 2024; Schultner, Stillerman et al., 2024). These instrumental preferences, expressed in choice behaviors, have been found to predict subsequent social decisions independently of self-reported preferences and IAT measures of implicit evaluation (Hackel et al., 2022b; Traast et al., 2024). Consistent with models of instrumental learning, this form of implicit evaluation operates implicitly and is expressed most directly in goal-directed behavior—features that align it with motivation-oriented theories of social cognition (Ferguson & Bargh, 2004; Strack & Deutsch, 2004).

Together, these findings clarify that "implicit evaluation" can refer to different underlying memory systems—semantic, affective (i.e., 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 (Figure 3), and that theories of implicit evaluation built only on models of semantic memory and data from conventional implicit tasks are incomplete.

Traits vs. evaluations. Traits and evaluations have long been distinguished in both impression formation and intergroup bias (Asch, 1946; Allport, 1954, Amodio & Ratner, 2011; Carlston, 1994; Devine, 1989; Dovidio et al., 1996). Traits, like stereotypes, refer to person or group characteristics and are represented as beliefs and conceptual associations in semantic memory. Evaluations, by contrast, refer to a perceiver's preference toward 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, studies using unconfound assessments, in which measures of evaluation do not include stereotypes and, conversely, measures of stereotyping are equated on valence, observed weak correlations between stereotyping and evaluation (Amodio & Devine, 2006; Amodio & Hamilton, 2012; Bijlstra et al., 2010; Dovidio et al., 2004; Gilbert et al., 2012; Glaser & Knowles, 2008; Wittenbrink et al., 1997, 2001). Research on spontaneous impression formation has similarly observed dissociations in the formation and effects of trait and evaluative inferences (Schneid et al., 2015; Olcaysoy Okten et al., 2019).

It is notable that this trait-evaluation distinction differs from the position that stereotypes and prejudice emerge from a single underlying representation (Kurdi et al., 2019). However, evidence for the single-representation position has come from measures or manipulations that confound stereotype traits with valence (Kurdi et al., 2019; Phills et al., 2020). For example, Kurdi et al. reported large correlations between IAT measures of implicit prejudice and stereotyping when stereotypes with positive and negative valence were used. However, when they used unconfounded IAT measures of prejudice and stereotyping, the intercorrelations were small and similar in effect size to prior work supporting a stereotype-evaluation distinction (e.g., Amodio & Devine, 2006; Gilbert et al., 2012).

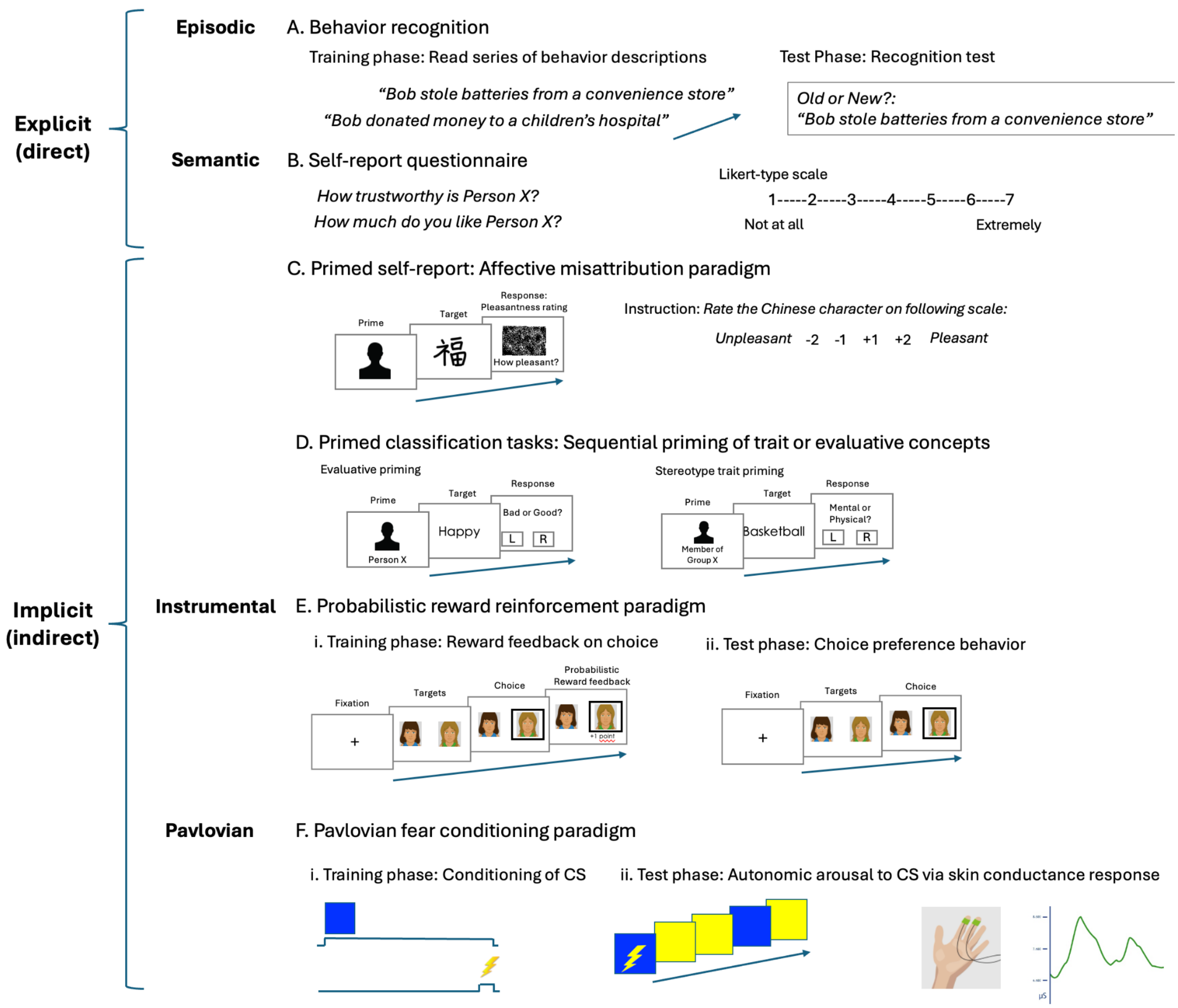


Figure 3. Experimental paradigms for assessing impression formation as represented by different learning and memory mechanisms, including both explicit (direct) and implicit (indirect) assessments.

Independent effects on expression. A key contribution of a learning and memory framework is that it predicts how impressions are expressed in behavior (see Fig 2). Whereas semantic and episodic impressions guide our explicit thoughts, judgments, and plans regarding a person, instrumental associations implicitly guide behavior in decisions and social interactions. Habits guide automatic actions to previously-rewarded cues, whereas Pavlovian associations guide responses to potential threats through freezing, attentional vigilance, and physiological readiness (i.e., conditioned suppression; Reiter & DeVellis, 1976; Roelofs et al., 2010).

In an early demonstration of these effects, Amodio and Devine (2006) showed that White Americans' scores on an IAT measure of implicit stereotypes uniquely predicted their trait impressions of a Black partner, whereas scores on an implicit prejudice IAT, proposed at the time to reflect an affective Pavlovian association, uniquely predicted their seating distance from a Black partner (Amodio & Devine, 2006). In other research, feelings of intergroup anxiety, associated with a Pavlovian response, selectively enhanced the expression of implicit prejudice but not implicit stereotypes (Amodio & Hamilton, 2012). These patterns resemble previous dissociations between explicit cognitive and affective measures of intergroup bias (Dovidio et al., 1996, 2004) and between effects of explicit prejudice beliefs and implicit race evaluations on interracial interaction behavior (Dovidio et al., 1997, 2002; Fazio et al., 1995).

Research has also distinguished the effects of trait-based and reward-based impressions on participants' social decisions (Hackel et al., 2015). Whereas instrumental reward associations tend to be more strongly expressed in behavioral choices to interact with partners, semantic trait associations are more strongly expressed in self-reported social preferences and intentions for future interaction (Hackel et al., 2015; 2020; Traast et al., 2024). In other work, impressions based in episodic and semantic knowledge were shown to play different roles in decisions to help somone, based either on recalling the exact amount that person had donated to a charity (episodic) or a gist description of the donation as "some" or "none" (semantic) (Hackel & Mende-Siedlecki, 2023). The dissociation between semantic and episodic aspects of an impression was also shown using a directed forgetting procedure: although instructions to forget a behavior associated with a face impaired later episodic memory for the behavior, the trait implied by the behavior remained semantically accessible and continued to influence person judgment (Hupbach, Olcaysoy Okten, & Horn, 2022).

In cases where two or more memory systems compete to influence a response, the expression of one over another may be moderated by situational factors. For example, although episodic and instrumental learning normally function in concert, distraction (i.e., cognitive load) selectively impairs episodic memory, leaving instrumentally-learned responses to primarily drive performance (Foerde et al., 2006). 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 (Schneid et al., 2015).

The timing and certainty of information during learning can also affect the expression of competing memory systems. Studies of feedback-based learning show that when feedback is immediate, humans simultaneously form episodic and instrumental associations, but when feedback is delayed by even a few seconds, instrumental learning is selectively impaired and only episodic learning occurs (Foerde & Shohamy, 2011; Foerde et al., 2013). Similarly, in uncertain environments, one relies more on episodic memory than on instrumental associations in decision making, consistent with a shift from automatic to deliberative processing (Daw et al., 2005; Nicholas et al., 2022). These findings likely have implications for impression formation in situations marked by feedback delay or uncertainty, such as in online communication.

This research reveals that different components of an impression (e.g., semantic, episodic, instrumental, and Pavlovian) are expressed in different response channels, and that their expression may be moderated by specific situational factors.

Interactive effects. Memory systems also function interactively, such that they can shape each other's operation (Poldrack & Packard, 2003). While such interactions have been demonstrated extensively in nonsocial domains (Doll et al., 2009; Foerde et al., 2006; Lindström et al., 2019; Phelps, 2004), they are likely to have similar effects in social contexts (Amodio, 2019).

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 (Kensinger, 2009; LaBar & Phelps, 1998; McGaugh, 2004). Although much prior work has examined mood effects on impression formation (e.g., effects of sad vs. happy mood) (Forgas, 2020), this Pavlovian-episodic memory system interaction suggests that fear-based arousal in particular should enhance the encoding of episodic person memory—a prediction consistent with observations of a negative bias in impression formation (Skowronski & Carlston, 1989).

My colleagues and I recently examined the interactive effect of semantic and instrumental systems in prejudice formation (Schultner, Stillerman et al., 2024). We proposed that mere knowledge of a societal stereotypes, a form of semantic memory, can bias how a perceiver experiences and learns from members of the stereotyped group through instrumental learning in subsequent social interactions, leading to the internalization of prejudice. This memory systems interaction— between declarative semantic knowledge and a nondeclarative instrumental learning process—describes a process through which exposure to societal factors can transform into individual-level implicit attitudes (Rösler et al., 2024; Traast et al., 2024).

Semantic knowledge, such as a preexisting preferences or stereotypes, may also prevent individual from engaging in instrumental social-interactive learning. For example, if a person holds a positive impression of a particular group, they may selectively interact with its members and thus never form or update impressions of other groups (Denrell, 2005; Fazio et al., 2004). This selective exposure effect has been proposed as a mechanism through which group prejudices and stereotypes are formed and reinforced (Allidina & Cunningham, 2021; Bai et al., 2022; Fazio et al., 2004).

Although research has just begun to directly 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.

Section summary. 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 behavior.

Impression updating

Will Rogers famously quipped, “You never get a second chance to make a first impression.” From a learning and memory perspective, 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 (Asch, 1946; Cone et al., 2015). This focus on evaluative updating, as opposed to trait updating, may 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. Evaluations, by contrast, may 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 program of work, distinct patterns of trait and evaluative updating were demonstrated in the context of spontaneous trait and evaluative inferences. Prior findings showed that spontaneous trait and evaluative inferences comprise distinct representations, formed in parallel (Schneid et al., 2015); building on this work, Olcaysoy Okten et al. (2019) found that only spontaneous evaluative inferences were updated in response to new impression-inconsistent information about a target’s behavior, consistent with an instrumental learning process. The updating of spontaneous trait inferences, by contrast, involved the encoding of new traits alongside the old traits, consistent with a basis in semantic memory. Subsequent work has shown that spontaneous trait inference updating does not involve the replacement of old traits, but rather the addition of new trait information, and that this new information is selected during social judgments (Olcaysoy Okten & Moskowitz, 2024).

Research on the instrumental learning of impressions has used computational reinforcement learning models to demonstrate updating (Hackel & Amodio, 2018; Lockwood & Klein-Flügge, 2021). Consistent with reinforcement learning theory (Sutton & Barto, 1998), these models specify the incremental, trial-by-trial updating of a reward association (i.e., expected value) in response to new information. By showing that behavioral data from instrumental impression formation tasks fit best to such models, these studies provide strong evidence for an instrumental learning mechanism of updating (Schultner et al., 2024; Traast et al., 2024).

Given the different expressions of memory systems in behavior, an assessment of updating must be sensitive to the underlying representation of interest. Measures that rely on self-report, which include questionnaires and some implicit tasks such as the AMP, are be primarily sensitive to changes in semantic learning. Measures that rely on action (e.g., behavioral classifications) and feedback, like most reinforcement learning paradigms (e.g., probabilistic selection tasks), are primarily sensitive to changes in instrumental learning. Behavioral tasks that pick up on freezing or response slowing are sensitive to changes in Pavlovian threat associations. To the extent a task combines these response features (e.g., as in the IAT, evaluative priming, some versions of the AMP), it may be sensitive to multiple underlying memory processes. If a measure mismatches the underlying learning process, then updating effects may be obscured.

A consideration of mechanism-measure match may illuminate longstanding questions about the nature of impression updating, such as whether implicit impression updating occurs slowly or rapidly (Rydell & McConnel, 2006; Cone et al., 2015). In experiments by Rydell and McConnell, participants formed impressions of a target person by reading statements about a behavior, 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. The authors found that a change in the valence of target behaviors produced a rapid change in evaluation on a self-report measure but a relatively slow change on the IAT, an implicit task that involves behavioral choice classifications. By comparison, Cone et al. (2015, 2017) used a similar impression formation task and found that, in response to a single extreme countervailing behavior, both implicit and explicit evaluations were updated. However, in their studies, implicit evaluation was measured using the AMP, a task in which participants make evaluative self-report judgments of targets following a positive or negative prime (Payne et al., 2005). Thus, it is possible that the discrepancy between findings reflects the different implicit measures: whereas responses on both the AMP and IAT involve a combination of semantic and instrumental processes, the AMP's greater sensitivity to semantic knowledge, relative to the IAT, should reveal more dramatic updating. Indeed, when the Cone et al. procedure was used with an IAT measure of updating, the signature reversal in impression valence was not observed (Cone & Calanchini, 2020). This analysis highlights the utility of a memory systems analysis for interpreting patterns of impression updating, and it provides a bases for developing interventions for impression change.

VI. 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 advances these debates by considering contemporary memory research that extends beyond conventional models of social cognition.

What exactly are implicit impressions? And can they be nonconscious?

Few topics in social cognition spark as much debate as the nature and significance of implicit processes and the tasks designed to measure them (Gawronski et al., 2022). Are implicit processes truly nonconscious? Unintentional? Or merely indirect? And how do they relate to behavior? Many scholars have called for greater clarity in defining the construct (Melnikoff & Bargh, 2018; Gawronski et al., 2022), while others suggest abandoning it altogether (Corneille & Hutter, 2020). From the perspective of learning and memory, however, these debates partly stem 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 is incorporates studies of nonhuman animals (e.g., mice, sea slugs)—subjects that cannot self-report and may lack the capacity for semantic cognition. These studies necessitated the development of models of learning and behavior, such as Pavlovian conditioning and instrumental learning, that do not rely on explicit reports or semantic processes. Furthermore, research using animal models permits the identification of neural circuits underlying these implicit forms of learning and behavior, enabling these processes to be theoretically distinguished from other cognitive mechanisms.

Studies of human brain lesion patients further elucidated the nature of implicit memory processes. Research on temporal lobe patients, such as H.M., demonstrates that implicit associations, involving Pavlovian or instrumental learning, can occur without declarative knowledge of what was learned (Bechara et al., 1995; Knowlton et al., 1996; LaBar & Phelps, 1998). Conversely, patients 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. fMRI studies of healthy individuals provide additional insight, showing that while the neural substrates of implicit and explicit processes are dissociable, they frequently co-occur and may create the appearance of a unified response in beahvior (Foerde et al., 2006). This approach reveals that implicit and explicit processes involve the coordinated activity of multiple memory systems rather than a single (e.g., propositional) mechanism.

This body of evidence has inspired a range of experimental tasks and methods designed to isolate mulitple 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 may limit the measurement and interpretation of implicit or nonconscious processes.

What does learning and memory research tell us about the measurement of implicit impressions? Some forms of memory—episodic and semantic—are declarative (reportable), can be expressed directly (explicitly), and are typically subject to awareness (Figure 4). Thus, semantic and episodic associations can be assessed using either explicit measures, such as self-reports, although they may also be observed indirectly in implicit tasks that assess conceptual associations, such as semantic priming. Other forms of memory—instrumental (including habits) and Pavlovian—are nondeclarative, expressed indirectly, and may operate outside of conscious awareness. As such, instrumental and Pavlovian associations can only be observed with indirect (implicit) measures, such as probabilistic classification or fear conditioning tasks, as these associations are not directly accessible to awareness and thus not reportable.

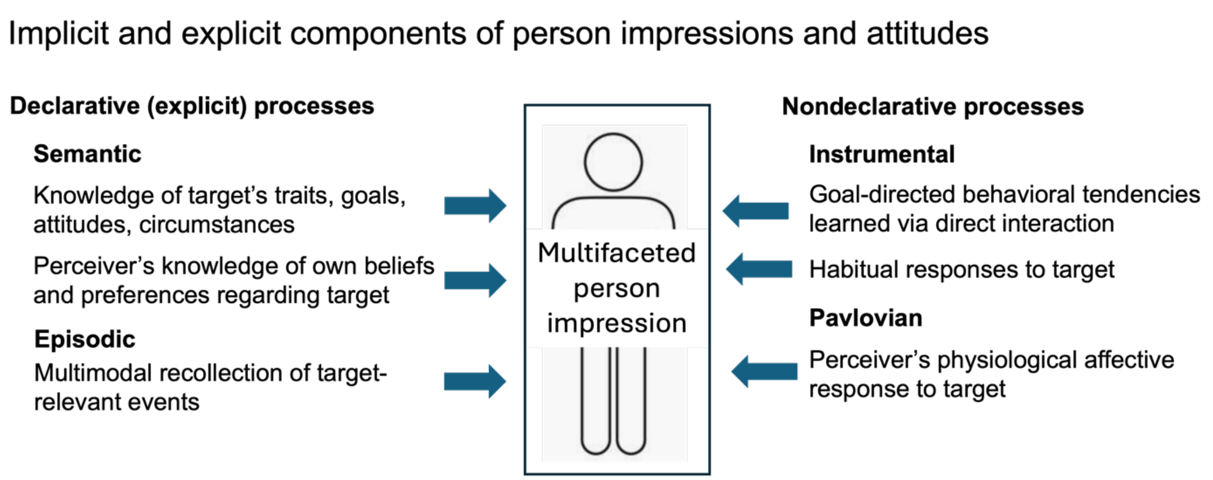


Figure 4. Implicit and explicit components of impressions can be described in terms of declarative and nondeclarative learning and memory systems.

It is notable that most implicit social cognition tasks, such as the IAT, AMP, and semantic/evaluative priming tasks, blur these distinctions between memory systems. That is, they measure semantic associations (i.e., of traits or evaluations) with an indirect assessment. While such tasks may give the appearance of a nonconscious semantic association, a learning and memory analysis suggests such measures capture the indirect expression of declarative (i.e., conscious) knowledge. This interpretation is consistent with evidence that people are often aware of associations expressed on implicit tasks (De Houwer, 2006; Gregg et al., 2006; Hahn et al., 2014; Morris & Kurdi, 2023).

To measure nonconscious associations, methods tailored to nondeclarative memory processes are required. For instance, instrumental learning may be assessed using tasks that afford the formation of action-reward associations while hindering semantic learning, such as probabilistic selection tasks (Frank et al., 2004; Knowlton et al., 1996). Studies of impression formation have adapted these tasks to demonstrate implicit social preferences that are independent of participants' subjective attitudes or semantic associations (Hackel et al., 2015; Schultner, Stillerman et al., 2024; Traast et al., 2024). Similarly, Pavlovian learning can be assessed using Pavlovian fear conditioning paradigms paired with measures of freezing or autonomic arousal (Bechara et al., 1999). Habits may be assessed using reward devaluation tasks which measure learned behaviors that persist after they are no longer goal-relevant (Foerde, 2018). Critically, some tasks engage a combination of memory processes, whereas others may assess only one component of a multi-system response. Careful task design and interpretation are thus crucial for isolating and understanding impression representations of interest.

It has been noted that while an actor may be aware of possessing a belief or association, they may be unaware of its expression in behavior or the processes through which it is expressed (Gawronski et al., 2022). For example, a math professor might be aware of his gender stereotype beliefs but unaware of how they influence his grading decisions. In studies of implicit impressions, a participant may 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 of semantic and instrumental processes in most implicit tasks; that is, while one's belief is represented in semantic memory, which is declarative, its influence on task behavior, which involves target classifications in semantic priming, relies on an instrumental process that is nondeclarative (Schultner, Stillerman et al., 2024; Solarz, 1960). Thus, a memory systems analysis clarifies why some aspects of an implicit response are subject to awareness while others are not.

Single- vs. dual vs. multiprocess accounts of impressions and attitudes

How many processes are needed to explain impression formation? If we assume that impressions are based on known mechanisms of learning and memory, then a multi-process account based on these memory systems is most plausible. This multi-process account, grounded in the functions and neural substrates of learning and memory, offers a deductive, model-based approach to predictions about impression formation. From this perspective, the critical issue is not the number of processes involved, but rather their specific functions in social cognition (Amodio, 2019; Henke, 2010).

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 processes: one that is associative, automatic, impulsive, and nonconscious and one that is propositional, deliberative, reflective, and conscious (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004; Smith & DeCoster, 2000). 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 evidence that single instances of explicit information can induce or change implicit evaluations (Gregg et al., 2003; DeHouwer, 2006; Kurdi & Dunham, 2020) and that participants are often aware of associations assessed by implicit tasks (Hahn et al., 2014; Kurdi & Morris, 2023).

From a learning and memory perspective, existing dual- and single-process models of social cognition can both 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 to only 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 behavior, such as those guided by instrumental responses, habits, episodes, or Pavlovian reactions, then neither single- nor dual-process accounts are sufficient.

It is notable that some dual process models describe associative processes as affective or motivational—features that intuitively correspond to Pavlovian or instrumental processes (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). An implication is that such models may benefit from respecification that includes more than one memory system. Doing so could expand their explanatory power and add precision to their measurement.

Summary and future directions (700 words)

The field of social cognition was borne of the insight that impression formation processes are rooted in learning and memory, originally importing theories of semantic memory to the study of person perception (Hastie et al., 1980). The current analysis continues this tradition by describing an updated perspective of impression formation informed by contemporary models of learning and memory. What are the major contributions of this updated approach?

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

Clarifying and expanding measurement. A learning and memory perspective acknowledges that conventional measures of social impressions and attitudes pertain primarily to semantic memory, which may limit their ability to assess aspects of impressions involving other forms of learning and memory. A learning and memory perspective suggests new methods for assessing a broader range of impressions and attitudes, along with a theoretical framework for interpreting them.

Addressing existing theoretical debates. A memory systems framework clarifies the role of awareness and implicit processes in impression formation, and it addresses the single- vs. dual-process debate by contextualizing it within a broader memory systems framework. 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 may be expressed in different kinds of responses.

Predicting behavior. A longstanding critique of impression formation research is that its measures often fail to predict behavior. While existing models typically focus on the formation, representation, and activation of impressions within the mind, a memory systems framework generates predictions for behavior based on known neurocognitive pathways through which memory influences decision and action. As such, it extends the scope of existing theories to explain how impressions are expressed in behavior.

What's next? With an updated framework in place, the next step is to more thoroughly test its predictions. Recent research has begun to explore interactive effects of semantic and instrumental processes to understand how stereotypes influence impression formation in direct social interactions (Schultner et al., 2024; Traast et al., 2025). Other research has used this approach to examine the unique roles of episodic memory and habits in impression formation (Hackel et al., 2019; Hackel & Mende-Siedlecki, 2023). Although earlier research has studied Pavlovian fear conditioning effects in group-based impressions (Amodio et al., 2003; Olsson et al., 2005), new 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 more fully explain how impressions function in dynamic social relationships (Box 2).

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 endeavor by specifying how individuals encode and internalize information from higher-level social structures and communicate it to others (Schultner, Stillerman et al., 2024; Schultner, Lindström et al., 2024). 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 toward these goals will require theoretical and methodological expertise that is increasingly interdisciplinary, for example, by incorporating theory and methods from cognitive neuroscience, computational cognition, and sociology into the social cognition curriculum. At the same time, as impression formation research builds interdisciplinary connections, its utility as a hub domain for examining the high-level functions of more basic cognitive processes is increasingly recognized by the broader field.

Conclusions

The study of impression formation and updating, at its core, concerns the way we learn about and remember people. By considering the multiple ways we can learn about people, the framework presented here advances our understanding of how impressions are formed, represented in the mind, expressed in behavior, and potentially changed. It provides an update to classic theories of social cognition, which were originally informed by studies of semantic memory, and brings us closer to Asch's (1946) holistic conceptualization of impression formation as a core capacity of the human mind.

Box 1. Memory systems & intergroup social cognition

Research on impression formation and intergroup social cognition are closely related: 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 our understanding of stereotyping and prejudice (Amodio & Cikara, 2021).

Clarifying representations of intergroup bias. From a memory systems perspective, and in line with classic theories (Hamilton & Sherman, 1996), stereotypes reflect knowledge in semantic memory and thus may be expressed directly in verbal reports and also indirectly in conceptual word classifications. Departing from classic theories, however, a memory systems analysis identifies multiple forms of prejudice (i.e., a group-level attitude). These correspond to semantic evaluation, instrumental reward associations, Pavlovian threat responses, and habit—each of which reflects a form of group-based preference. Because an individual's intergroup bias could involve any combination of these processes, this model accounts for why prejudiced attitudes do not always align with stereotype knowledge and why some forms of intergroup bias are more likely to produce discriminatory behavior (Amodio & Devine, 2006; Dovidio et al., 1996, 2002).

Measurement of intergroup bias. A memory systems model 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 (i.e., instrumental) processes. It further suggests that self-report and implicit measures may differ not merely because of the different response format (Payne et al., 2008), but because they afford expressions of different underlying processes. Intergroup researchers have been at the forefront of developing new methods to tap into these different underlying components of bias (Amodio et al., 2003; Dunsmoor et al., 2016; Hackel et al., 2020b; Phelps et al., 2000).

Predicting bias in behavior. A common critique of intergroup research is that implicit bias measures are often weakly associated with behavior. The memory systems approach was developed, in part, to clarify how measures of implicit bias should predict behavior (Ratner & Amodio, 2011). It suggests that measures of bias reflecting semantic associations should primarily predict high-level judgments and verbal behavior, whereas measures reflecting instrumental or Pavlovian associations should be more predictive of nonverbal behaviors (e.g., approach or social distancing).

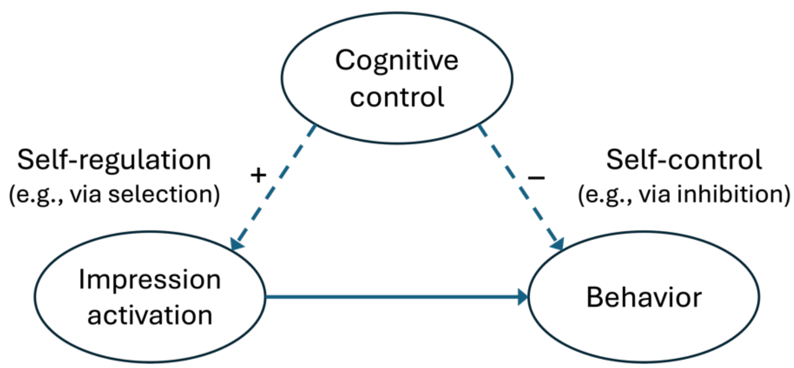
Prejudice reduction. A memory systems analysis informs prejudice reduction by clarifying which aspects of bias are changeable and how they may be changed. By considering mechanisms of updating, it informs the design of interventions assessment of their impacts. Furthermore, by identifying forms of prejudice that are difficult to change (e.g., Pavlovian associations, habits), this analysis highlights the importance of structural interventions that supersede individual-level responses (Rösler & Amodio, 2022). That is, it suggests that while models of individual-level processes are critical for understanding bias formation and expression, effective bias reduction often requires structural-level interventions.

Box 2. How are impressions regulated?

To serve us adaptively, our impressions often require regulation. When selecting a doctor, we'd do well to focus on their medical skills rather than their taste in fashion, just as a successful holiday dinner may require us to overlook uncle Bill's taste for fringe politics. These adjustments to impressions rely on cognitive control.

Classic dual process theories address how the activation of social knowledge—such as traits or stereotypes—is modulated through the engagement of control (Bodenhausen & Macrae, 2000; Devine, 1989; Fazio, 1990; Gilbert et al., 1998; Smith & DeCoster, 2000). Built on cognitive theories of information search in semantic memory (e.g., Shiffrin & Schneider, 1977), they conceptualize control as operating on the activation of an association in the mind; that is, by inhibiting a mental concept or overriding it with an alternative.

By comparison, cognitive neuroscience models, which incorporate studies of both human and nonhuman animals to address a broader range of responses, place greater emphasis on behavior (i.e., motor processes) as the primary target of inhibitory control (Aron et al., 2014; Badre, 2025; Miller & Cohen, 2001). That is, they suggest that cognitive control processes broadly support self-regulation broadly by coordinating the goal-directed operation of memory systems, but when self-control is needed (i.e., favoring one response over another), it occurs through inhibition or switching of actions.



Box 2 Fig. Cognitive control processes modulate impressions in two ways: by selectively activating goal-relevant aspects of an impression in the mind (i.e., *self-regulation*), and by inhibiting the influence of an impression on behavior (i.e., *self-control*).

Thus, When it comes to overriding a response (i.e., self-control), this analysis suggests that strategies targeting behavior are more effective than strategies targeting mental associations. This perspective aligns with evidence supporting the effectiveness of behavioral implementation intentions (Gollwitzer, 1999) and the ineffectiveness of mental suppression (Monteith et al., 1998; Wegner, 1994).

References

Allport, G. W. (1954). The nature of prejudice. Addison-Wesley.

Allidina, S., & Cunningham, W. A. (2021). Avoidance begets avoidance: A computational account of negative stereotype persistence. *Journal of Experimental Psychology: General*, *150*(10), 2078.

Ames, D. R. (2004). Inside the mind reader's tool kit: projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87*, 340-353.

Amodio, D. M. (2014). Dual experiences, multiple processes: Looking beyond dualities for mechanisms of the mind. In J. S. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 560-576). New York: Guilford Press.

Amodio, D. M. (2019). Social cognition 2.0: An interactive memory systems account. *Trends in Cognitive Sciences*, *23*(1), 21-33.

Amodio, D. M., & Cikara, M. (2021). The social neuroscience of prejudice. Annual Review of Psychology, 72, 439-469.

Amodio, D. M. & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. Journal of Personality and Social Psychology, 91, 652-661.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink responses and self-report. Journal of Personality and Social Psychology, 84, 738-753.

Amodio, D. M., & Hamilton, H. K. (2012). Intergroup anxiety effects on implicit racial evaluation and stereotyping. *Emotion*, *12*(6), 1273-1280.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. Current Directions in Psychological Science, 20, 143-148.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of experimental psychology*, *70*(4), 394.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in cognitive sciences*, *18*(6), 279-292.

Arenson, S. J., Lannon, P. B., Offermann, L. R., & Kafton, A. (1982). The validity of attitude change by classical conditioning. *The Journal of Social Psychology*, *117*(2), 243-248.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences, 18*, 177-185.

Asch, S. E. (1946). Forming impressions of personality. *The journal of abnormal and social psychology*, *41*(3), 258.

Baddeley, A. (2001). The concept of episodic memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1345-1350.

Badre, D. (2024). Cognitive control. *Annual Review of Psychology, 76*, 167-195.

Bai, X., Fiske, S. T., & Griffiths, T. L. (2022). Globally inaccurate stereotypes can result from locally adaptive exploration. Psychological Science, 33(5), 671-684.

Balcetis, E. (2008). Where the motivation resides and self-deception hides: How motivated cognition accomplishes self-deception. *Social and Personality Psychology Compass*, *2*(1), 361-381.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4-5), 407-419.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: the influence of trait information presented outside of conscious awareness on impression formation. *Journal of personality and Social psychology*, *43*(3), 437.

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of neuroscience*, *19*(13), 5473-5481.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*(5227), 1115-1118.

Biernat, M., & Sesko, A. K. (2013). Communicating about others: Motivations and consequences of race-based impressions. *Journal of Experimental Social Psychology*, *49*(1), 138-143.

Blaison, C., Imhoff, R., Hühnel, I., Hess, U., & Banse, R. (2012). The affect misattribution procedure: Hot or not?. *Emotion*, *12*(2), 403.

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, *46*(4), 657-663.

Binney, R. J., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience & Biobehavioral Reviews*, *112*, 28-38.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology, 51*, 93-120.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*(1), 15958.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological bulletin*, *114*(1), 80.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164-169.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of personality and social psychology*, *47*(6), 1191.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Skrull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35*, 38–48.

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., & Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature communications*, *11*(1), 2100.

Carlston, D. E. (1994). *Associated systems theory: A systematic approach to cognitive representations of persons*. Lawrence Erlbaum Associates, Inc.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840.

Chen, J. M. (2019). An integrative review of impression formation processes for multiracial individuals. *Social and Personality Psychology Compass, 13*, e12430.

Cho, H., & Hackel, L. M. (2022). Instrumental learning of social affiliation through outcome and intention. Journal of Experimental Psychology: General.

Clore, G. L., & Gormly, J. B. (1974). Knowing, feeling, and liking a psychophysiological study of attraction. *Journal of Research in Personality*, *8*(3), 218-230.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Cone, J., & Calanchini, J. (2021). A process dissociation model of implicit rapid revision in response to diagnostic revelations. *Personality and Social Psychology Bulletin*, *47*(2), 201-215.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37.

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In *Advances in experimental social psychology* (Vol. 56, pp. 131-199). Academic Press.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. Social cognitive and affective neuroscience, 7(7), 764-770.

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, *24*(3), 212-232.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: a sociofunctional threat-based approach to" prejudice". *Journal of personality and social psychology*, *88*(5), 770-789.

Cushman, F. (2024). Computational social psychology. *Annual Review of Psychology*, *75*, 625-652.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, *8*(12), 1704-1711.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. Learning and Motivation, 37(2), 176-187.

De Houwer, J., Hermans, D., and Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion, 12*, 145–169.

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biological psychology*, *73*(1), 39-48.

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological review*, *112*(4), 951.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, *56*(1), 5-18.

Dickerson, B. C., & Eichenbaum, H. (2010). The episodic memory system: neurocircuitry and disorders. Neuropsychopharmacology, 35(1), 86-104.

Dodson, C. S., Darragh, J., & Williams, A. (2008). Stereotypes and retrieval-provoked illusory source recollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 460.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain research*, *1299*, 74-94.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice and discrimination: Another look. In C. N. McCrae, C. Stangor, & M. Hewstone (Eds.), (pp. 276 –319). New York: Guilford.

Dovidio, J. F., Esses, V. M., Beach, K. R., & Gaertner, S. L. (2004). The role of affect in determining intergroup behavior: The case of willingness to engage in intergroup affect. In D. M. Mackie & E. R. Smith (Eds.), From prejudice to intergroup emotions: Differentiated reactions to social groups (pp. 153–171). Philadelphia: Psychology Press.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, *82*(1), 62-68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of experimental social psychology*, *33*(5), 510-540.

Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, *39*(5), 420-433.

Dunsmoor, J. E., Kubota, J. T., Li, J., Coelho, C. A., & Phelps, E. A. (2016). Racial stereotypes impair flexibility of emotional learning. *Social cognitive and affective neuroscience*, *11*(9), 1363-1373.

Elsey, J. W., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. Psychological bulletin, 144(8), 797-848.

Enge, L. R., Lupo, A. K., & Zárate, M. A. (2015). Neurocognitive Mechanisms of Prejudice Formation: The Role of Time-Dependent Memory Consolidation. *Psychological Science, 26(7)*, 964-971. https://doi.org/10.1177/0956797615572903

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 75-109). Academic Press.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of personality and social psychology*, *87*(3), 293.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?. *Journal of personality and social psychology*, *69*(6), 1013-1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of personality and social psychology*, *50*(2), 229-238.

FeldmanHall, O., Montez, D. F., Phelps, E. A., Davachi, L., & Murty, V. P. (2021). Hippocampus guides adaptive learning during dynamic social interactions. *Journal of Neuroscience*, *41*(6), 1340-1348.

FeldmanHall, O., & Nassar, M. R. (2021). The computational challenge of social learning. Trends in Cognitive Sciences, 25(12), 1045-1057.

Fendt, M., & Fanselow, M. S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience & Biobehavioral Reviews*, *23*(5), 743-760.

Ferguson, M. J., & Bargh, J. A. (2004). Liking Is for Doing: The Effects of Goal Pursuit on Automatic Evaluation. *Journal of Personality and Social Psychology, 87*, 557–572. https://doi.org/10.1037/0022-3514.87.5.557

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 23, pp. 1–74). New York, NY: Academic Press.

Foerde, K. (2018). What are habits and do they depend on the striatum? A view from the study of neuropsychological populations. *Current opinion in behavioral sciences*, *20*, 17-24.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *proceedings of the National Academy of Sciences*, *103*(31), 11778-11783.

Foerde, K., Race, E., Verfaellie, M., & Shohamy, D. (2013). A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. *Journal of Neuroscience*, *33*(13), 5698-5704.

Foerde, K., & Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience*, *31*(37), 13157-13167.

Forgas, J. P. (2020). Affect and social judgments: An introductory review. *Emotion and social judgements*, 3-29.

Förster, J., Liberman, N., & Friedman, R.S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs Personality and Social Psychology Review, 11, 211-233.

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940-1943.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological review*, *118*(2), 247.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual review of psychology*, *49*(1), 87-115.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological bulletin*, *132*(5), 692.

Gawronski, B., Ledgerwood, A., & Eastwick, P. W. (2022). Implicit bias≠ bias on implicit measures. *Psychological Inquiry, 33*, 139-155.

Gilbert, D. T., Krull, D. S., & Pelham, B. W. (1988). Of thoughts unspoken: Social inference and the self-regulation of behavior. *Journal of Personality and Social Psychology, 55*, 685–694.

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*, *50*(14), 3600-3611.

Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, *44*(1), 164-172.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. American Psychologist, 54(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Graybiel, A. M., & Grafton, S. T. (2015). The striatum: where skills and habits meet. *Cold Spring Harbor perspectives in biology*, *7*(8), a021691.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychological review, 102(1), 4.

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. American Psychologist, 72(9), 861-871.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. Journal of Personality and Social Psychology, 90(1), 1–20. https://doi.org/10.1037/0022-3514.90.1.1

Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. Current Opinion in Psychology, 24, 92-97.

Hackel, L. M., Berg, J. J., Lindström, B. R., & Amodio, D. M. (2019). Model-Based and Model-Free Social Cognition: Investigating the role of habit in social attitude formation and choice. Frontiers in Psychology, 10, article 2592. doi: 10.3389/fpsyg.2019.02592

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on decision making. Nature Neuroscience, 18, 1233-1235.

Hackel, L. M., Kogon, D., Amodio, D. M., & Wood, W. (2022). Group value learned through interactions with members: A reinforcement learning account. Journal of Experimental Social Psychology, 99, 104267.

Hackel, L. M., & Mende-Siedlecki, P. (2023). Two modes of social impressions and their effects on choice. *Journal of Experimental Psychology: General, 152*(11), 3002–3020. https://doi.org/10.1037/xge0001432

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. Journal of Experimental Social Psychology, 88, 103948.

Hackel, L. M., Mende-Siedlecki, P., Loken, S., & Amodio, D. M. (2022). Context-dependent learning in social interaction: Trait impressions support flexible social choices. *Journal of Personality and Social Psychology, 123*, 655-675. https://doi.org/10.31234/osf.io/symrj

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369-1392.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological review*, *103*(2), 336.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. Journal of Experimental Social Psychology, 41, 129-140.

Hastie, R., Ostrom, T., Ebbesen, E., Wyer, R., Hamilton, D., & Carlston, D. (1980). *Person memory: Cognitive basis of social perception*. Hillsdale, N.J.: Erlbaum Associates.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological review*, *93*(3), 258.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological review*, *51*(6), 358.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*(7), 523-532.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of experimental social psychology*, *13*(2), 141-154.
Hull, C. L. (1943). Principles of behavior: an introduction to behavior theory.
Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: updating or source confusion?. *Memory, 17*, 502-510.
Hupbach, A., Olcaysoy Okten, I., & Horn, P. (2022). Directed forgetting in the social domain: Forgetting behaviors but not inferred traits. *Journal of Applied Research in Memory and Cognition, 11*, 522.
Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General*, *146*(7), 924.
Vatansever, D., Smallwood, J., & Jefferies, E. (2021). Varying demands for cognitive control reveals shared neural processes supporting semantic and episodic memory retrieval. Nature communications, 12(1), 2134.
John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: studies of trait use and accessibility in different contexts. *Journal of personality and social psychology*, *60*(3), 348.
Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology* (Vol. 2, pp. 219-266). Academic Press.
Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In *Advances in experimental social psychology* (Vol. 55, pp. 1-80). Academic Press.
Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1*, 99-113.
Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences*, *101*(9), 3310-3315.
Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: a test of extinction as updating mechanism. *Biological psychology*, *92*(1), 43-50.
Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: erasing human fear responses and preventing the return of fear. Nature neuroscience, 12(3), 256-258.
Klauer, K. C. (1997). Affective priming. *European review of social psychology*, *8*(1), 67-103.
Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in psychology*, *6*, 140545.
Kadwe, P. P., Sklenar, A. M., Frankenstein, A. N., Levy, P. U., & Leshikar, E. D. (2022). The influence of memory on approach and avoidance decisions: Investigating the role of episodic memory in social decision making. *Cognition*, *225*, 105072.
Klein, S. B., Cosmides, L., Gangi, C. E., Jackson, B., Tooby, J., & Costabile, K. A. (2009). Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Social Cognition*, *27*(2), 283-319.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399-1402..

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & memory*, *1*(2), 106-120.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40-80.

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition, 38*(Supplement), s42-s67.

Kruse, F., & Degner, J. (2021). Spontaneous state inferences. *Journal of Personality and Social Psychology, 121*, 774–791.

Kunda, Z., Sinclair, L., & Griffin, D. (1997). Equal ratings but separate meanings: Stereotypes and the construal of traits. *Journal of Personality and Social Psychology*, *72*(4), 720.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284–308. https://doi.org/10.1037/0033-295X.103.2.284

Kurdi, B., Mann, T. C., Charlesworth, T. E., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences*, *116*(13), 5862-5871.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science, 9*, 490-493.

Lieberman, M. D. (2000). Intuition: a social cognitive neuroscience approach. Psychological bulletin, 126, 109-137.

Liljeholm, M., & O'Doherty, J. P. (2012). Contributions of the striatum to learning, motivation, and performance: an associative account. *Trends in cognitive sciences*, *16*(9), 467-475.

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. Nature Communications, 12(1), 5168.

Lindström, B., Golkar, A., Jangard, S., Tobler, P. N., & Olsson, A. (2019). Social threat learning transfers to decision making in humans. *Proceedings of the National Academy of Sciences*, *116*(10), 4732-4737.

Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour—a reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761-771.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: the creation of new memories. *Journal of experimental psychology: General, 118*, 100.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*(3), 463.

Malle, B. F., & Holbrook, J. (2012). Is There a Hierarchy of Social Inferences? The Likelihood and Speed of Inferring Intentionality, Mind, and Personality. *Journal of Personality and Social Psychology, 102*, 661-684.

March, D. S., Gaertner, L., & Olson, M. A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, *29*(1), 1-13.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual review of neuroscience*, *24*(1), 897-931.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823.

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behavioral neuroscience*, *107*(1), 3.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci., 27*, 1-28.

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition*, *21*(2), 121-156.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. Trends in cognitive sciences, 22(4), 280-293.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual review of neuroscience, 24(1), 167-202.

Monfils, M. H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. science, 324(5929), 951-955.

Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review, 2*, 63-82.

Morris, A., & Kurdi, B. (2023). Awareness of implicit attitudes: Large-scale investigations of mechanism and scope. Journal of Experimental Psychology: General, 152(12), 3311–3343. https://doi.org/10.1037/xge0001464

Moskowitz, G. B., Okten, I. O., & Schneid, E. (2023). The updating of first impressions. In *The Handbook of Impression Formation* (pp. 348-392). Routledge.

Moskowitz, G.B. (2024). *Introduction to Social Cognition: The Essential Questions and Ideas*. New York, NY: Guilford Press.

Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology: General*, *145*(5), 548.

Nicholas, J., Daw, N. D., & Shohamy, D. (2022). Uncertainty alters the balance between incremental learning and episodic memory. *Elife*, *11*, e81679.

Nunes, L. D., Garcia-Marques, L., Ferreira, M. B., & Ramos, T. (2017). Inferential costs of trait centrality in impression formation: Organization in memory and misremembering. *Frontiers in Psychology*, *8*, 248652.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *science*, *304*(5669), 452-454.

Öhman, A. (1988). Nonconscious control of autonomic responses: A role for Pavlovian conditioning?. Biological Psychology, 27(2), 113-135.

Olson, I. R., McCoy, D., Klobusicky, E., & Ross, L. A. (2013). Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social cognitive and affective neuroscience*, *8*(2), 123-133.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. Psychological science, 12(5), 413-417.

Olcaysoy Okten, I., & Moskowitz, G. B. (2018). Goal vs. trait explanations: attributions beyond a person-situation dichotomy. *Journal of Personality and Social Psychology*, 114(2), 211-229.

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Easy to make, hard to revise: Updating spontaneous trait inferences in the presence of trait -inconsistent information. *Social Cognition, 38*(6), 571-625.

Olcaysoy Okten, I., Schneid, E., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology*. 117(1), 1-25. https://doi.org/10.1037/pspa0000156

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*(5735), 785-787.

Pavlov, I. P. (1927). Conditioned Reflexes. London: Oxford University Press.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. Journal of Personality and Social Psychology, 94(1), 16–31. https://doi.org/10.1037/0022-3514.94.1.16

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current opinion in neurobiology*, *14*(2), 198-202.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. Journal of cognitive neuroscience, 12(5), 729-738.

Phills, C. E., Hahn, A., & Gawronski, B. (2020). The bidirectional causal relation between implicit stereotypes and implicit prejudice. *Personality and Social Psychology Bulletin*, *46*(9), 1318-1330.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia, 41(3), 245-251.

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology, 58*, 1048–1061.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & memory*, *1*(4), 217-229.

Reiter, L. A., & DeVellis, B. M. (1976). Conditioned suppression in humans produced by a human stimulus. *Journal of Personality and Social Psychology*, *34*(2), 223-227.

Rescorla, R. A. (1988). Behavioral studies of Pavlovian conditioning. Annual review of neuroscience, 11(1), 329-352.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W.F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

Robbins, T. W., & Costa, R. M. (2017). Habits. *Current biology*, *27*(22), R1200-R1206.

Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 20160206.

Roelofs, K., Hagenaars, M. A., & Stins, J. (2010). Facing freeze: social threat induces bodily freeze in humans. *Psychological science*, *21*(11), 1575-1581.

Rohr, M., & Wentura, D. (2022). How emotion relates to language and cognition, seen through the Lens of evaluative priming paradigms. *Frontiers in Psychology*, *13*, 911068.

Rösler, I. K. & Amodio, D. M. (2022). Neural Basis of Prejudice and Prejudice Reduction. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 7, 1200-1208.

Rösler, I. K., Kerber, I., & Amodio, D. M. (March, 2024). Effects of moral stereotypes on the formation and persistence of group preferences. https://doi.org/10.31234/osf.io/vkgrx

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, *15*(8), 549-562.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of personality and social psychology*, *91*(6), 995.

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49-53.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. Neuron, 76(4), 677-694.

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*(5), 681-696.

Schultner, D. T., Stillerman, B. S., Lindström, B. R., Hackel, L. M., Hagen, D. R., Jostmann, N. B., & Amodio, D. M. (2024). Transmission of societal stereotypes to individual-level prejudice through instrumental learning. *Proceedings of the National Academy of Sciences, 121*, e2414518121.

Schultner, D. T., Lindström, B. R., Cikara, M., & Amodio, D. M. (2024). Transmission of social bias through observational learning. *Science Advances, 10,* eadk2030. 10.1126/sciadv.adk2030

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Shelton, J. N., & Richeson, J. A. (2006). Interracial interactions: A relational approach. *Advances in experimental social psychology*, *38*, 121-181.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70*, 1126–1141.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127–190.

Skinner, B. F. (1963). Operant behavior. *American psychologist*, *18*(8), 503.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. Psychological Bulletin, 105(1), 131–142.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review*, *4*(2), 108-131.

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of experimental psychology*, *59*(4), 239-245.

Spruyt, A., Hermans, D., Pandelaere, M., De Houwer, J., & Eelen, P. (2004). On the replicability of the affective priming effect in the pronunciation task. *Experimental Psychology*, *51*(2), 109-115.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171-177.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, *93*(24), 13515-13522.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social psychology*, *37*(10), 1660.

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology*, *57*(1), 37-40.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of social issues*, *41*(3), 157-175.

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature neuroscience*, *19*(6), 795-797.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2005). Social metacognitive judgments: The role of retrieval-induced forgetting in person memory and impressions. *Journal of Memory and Language*, *52*(4), 535-550.

Strack, F., & Deutsch, R. (2004). Reflection and impulse as determinants of conscious and unconscious motivation. *Social motivation: Conscious and unconscious processes*, 91-112.

Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135, pp. 223–260). MIT Press.

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194-199.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of personality and social psychology*, *36*(7), 778.

Thorndike, E. L. (1932). The fundamentals of learning.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Traast, I. J., Schultner, D. T, Doosje, B., & Amodio, D. M. (in press). Race effects on impression formation in social interaction: An instrumental learning account. Journal of Experimental Psychology: General.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239–257.

Tulving, E. (1985). How many memory systems are there?. *American psychologist*, *40*(4), 385.

Tulving, E. (2002). Episodic memory: From mind to brain. Annual review of psychology, 53(1), 1-25.

Uleman, J. S., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. E. Carlston (Ed.), Oxford handbook of social cognition (pp. 53–73). Oxford University Press.

Wegner, D. M. (1994). Ironic processes of mental control.Psychological Review, 101(1), 34–52. https://doi.org/10.1037/0033-295X.101.1.34

Wentura, D., & Degner, J. (2010). Automatic evaluation isn't that crude! Moderation of masked affective priming by type of valence. *Cognition & Emotion*, *24*(4), 609-628.

Wichert, S., Wolf, O. T., & Schwabe, L. (2013). Updating of episodic memories depends on the strength of new learning after memory reactivation. *Behavioral Neuroscience, 127*, 331-338.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of personality and social psychology*, *47*(2), 237.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of personality and social psychology*, *72*(2), 262-274.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of experimental social psychology*, *37*(3), 244-252.

Wood, W. (2017). Habit in personality and social psychology. *Personality and social psychology review*, *21*(4), 389-403.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. Psychological review, 114(4), 843.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of experimental social psychology*, *10*(2), 109-120.

Wyer Jr, R. S. (1980). The acquisition and use of social knowledge: Basic postulates and representative research. *Personality and Social Psychology Bulletin*, *6*(4), 558-573.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464-476.