Running head: Model-free social cognition

Model-Based and Model-Free Social Cognition

Leor M. Hackel[1], Jeffrey J. Berg[2], Björn R. Lindström[3], David M. Amodio[2,3]

[1]University of Southern California, [2]New York University, [3]University of Amsterdam

Please direct correspondence to:

Leor M. Hackel
Department of Psychology
University of Southern California
3620 South McClintock Avenue
Los Angeles, CA 90089
lhackel@usc.edu

or

David M. Amodio
Department of Psychology
New York University
6 Washington Place
New York, NY 10003, USA
david.amodio@nyu.edu

Abstract

Do habits play a role in our social impressions? To investigate the contribution of habits to the formation of social attitudes, we examined the roles of model-free and model-based reinforcement learning in social interactions—computations linked in past work to habit and planning, respectively. Participants in this study learned about novel individuals in a sequential reinforcement learning paradigm, choosing financial advisors who led them to high- or low-paying stocks. Results indicated that participants relied on both model-based and model-free learning, such that each independently predicted choice during the learning task and self-reported liking in a post-task assessment. Specifically, participants liked advisors who could provide large future rewards as well as advisors who had provided them with large rewards in the past. Moreover, participants varied in their use of model-based and model-free learning strategies, and this individual difference influenced the way in which learning related to self-reported attitudes: among participants who relied more on model-free learning, model-free social learning related more to post-task attitudes. We discuss implications for attitudes, trait impressions, and social behavior, as well as the role of habits in a memory systems model of social cognition.

Model-Based and Model-Free Social Cognition

Human thriving depends on social relationships, and the impressions we form of new acquaintances are essential guides to our social behavior. Not surprisingly, people are highly motivated to form impressions (Fitzsimons & Andersen, 2013): we might befriend people who are kind, hire people who are competent, avoid those who are domineering, or seek counsel from those who are empathic. That is, our impressions influence a range of goal-directed responses, whether initiated consciously or nonconsciously (Bargh & Ferguson, 2000; Brewer, 1988; Fiske & Neuberg, 1990; Uleman, 1999; Wyer & Srull, 1989). We use our knowledge of other people—of their traits, mental states, and behaviors—to predict their actions and decide whether to interact with them in light of our goals (Heider, 1958; Thornton & Tamir, 2018).

Yet, although social goals drive much of human social behavior, this is not always the case. Habits, in particular, are responses that occur automatically and independent of our goals, often representing a highly-repeated behavior that was once goal-directed but that persists and is expressed even when the goal has changed (Robbins & Costa, 2017; Wood & Rünger, 2016). Habits likely explain many behaviors, from benign compulsions like biting one's nails to more harmful acts like mindlessly reaching for a cigarette. Here, we asked whether habits may also contribute to social cognition—how we learn about, interact with, and evaluate other people—and thus help explain social behaviors that appear to occur independently of, or in opposition to, one's goals.

Multiple systems for social learning

Research on impression formation has, to date, primarily emphasized conceptual forms of learning that give rise to goal-directed behavior; that is, acquiring conceptual knowledge about a person's traits, behavior, and worth. For instance, people might associate a target person with concepts like "generous" or "competent." Early theories of impression formation focused on instructed forms of learning, in which we learn about a person from descriptions shared by others (Asch, 1946; Wyer & Carlston, 1979). If we are told that Bob is generous and friendly, we may infer

that he's a good person. We can also learn about other people through observation and the use of attributional processing (Heider, 1958; Jones & Davis, 1965; Rydell & McConnell, 2006). If we see Jane offer money to a homeless person, we may infer from her actions that she is generous; if we see Jane choose a high-performing stock, we may infer that she is competent. These conceptual associations can give rise to goal-directed behaviors, like choosing to spend time with someone generous or to hire someone competent.

More recent research has shown that social attitudes and impressions can also be formed through reward-based instrumental learning in direct social interaction—trial-and-error learning in which people make choices and receive feedback (Hackel, Doll, & Amodio, 2015). For instance, one might choose a lunch partner and experience rewards when they share their food, or one might hire a financial advisor and experience rewards when their advice pays off. Through this feedback, one can learn the reward value of an individual while also inferring aspects of their traits (Hackel et al., 2015). Unlike instructed and observational forms of learning, which are typically passive (e.g., reading about another person), instrumental learning is active: it concerns feedback from another person regarding one's own actions. If, on most days, Bob's greeting to Jane is met with a smile, he will associate reward with his behavior toward Jane in addition to inferring that she is friendly.

Instrumental learning thus represents a distinct mode of learning in social interactions relative to conceptual knowledge (Amodio, 2019). Instead of inferring other people's qualities in order to decide how to interact with them, instrumental learning involves learning how to interact with others through direct action and feedback. That is, in traditional impression formation approaches, Bob learns to interact with Jane because he infers she is friendly, and he wants to be around friendly people. In instrumental learning, Bob learns to interact with Jane because he previously did so and received rewarding outcomes, such as social rewards like smiles and compliments or material rewards like money and food. He may like Jane as a result of those rewards, rather than as a result of qualities he attributes to her. Thus, instrumental learning

directly informs how we should interact with others given the rewards they provide. In this way, preferences acquired through instrumental learning may be more directly tied to behavior.

A role for habits in social cognition?

Over time, instrumentally learned responses may also be automatized into habits (Robbins & Costa, 2017; Thorndike, 1911). Although people may initially perform an action deliberately to achieve a goal, rewards can "stamp in" an association between a stimulus or context and a response, such that people later perform the response automatically. In contrast to skills, which are goal-directed action routines triggered intentionally, habits reflect a well-learned response that unfolds even when it is not consistent with a goal, and it persists even when its expression is no longer rewarded (Balleine & Dickinson, 1998; Tricomi, Balleine, & O'Doherty, 2009; Wood & Rünger, 2016). Nevertheless, they can be adaptive, initiating an important behavior that we might otherwise forget in the pursuit of another goal, such as grabbing our keys when rushing out the door to get to work in the morning.

Habits differ from other forms of unintentional learning that may contribute to impression formation. For example, spontaneous trait impressions (STIs) form when a perceiver is simply asked to read and memorize a set of trait-implying sentences (Winter & Uleman, 1984; Carlston & Skowronski, 1994). People may be unaware that they formed an impression, yet STIs become evident in subsequent measures of cued recall and may subsequently influence judgment (Moskowitz & Roman, 1992). There is also evidence that evaluative conditioning, in which a neutral social target is paired repeatedly with either positive or negative images (Olson & Fazio, 2006; Walther, 2002), may even occur when such images are presented subliminally (e.g.., De Houwer, Hendrickx, & Baeyens, 1997; Hofmann et al., 2010; but see Sweldens, Cornielle, & Yzerbyt, 2014). However, both forms of learning involve passive exposure to stimuli and the formation of conceptual associations, likely supported by a semantic/conceptual associative memory system

(Amodio, 2019; Amodio & Berg, 2018), in contrast to the active process of action-outcome learning involved in instrumental habit formation.

Examining habit formation through reinforcement learning

A major challenge in the study of habits in humans is that it is often difficult to discern habits from other goal-directed processes in behavior. However, this distinction has recently been linked to two forms of behavior within a computational account of reinforcement learning (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). Broadly, reinforcement learning algorithms describe how an agent learns the value of different actions with different states of the world by making choices and experiencing rewards (Sutton & Barto, 1998). According to this account, two types of computations can underlie reinforcement learning: Agents can engage in *model-based learning,* in which they consider the likely outcomes of their actions given knowledge about their environment, and also in *model-free* learning, in which they associate actions directly with reward value and repeat previously rewarded actions (Daw et al., 2011; Doll, Duncan, Simon, Shohamy, & Daw, 2015). Model-based learning is thus prospective and goal-oriented, sensitive to both environmental contingencies (e.g., how to get to a reward) and expected outcomes (e.g., whether a desirable reward will be attained)—like a hungry mouse considering how to navigate a maze to reach the room with the tastiest cheese. In contrast, model-free learning is retrospective, relying on a past history of rewards for an action; it requires no internal model of one's environment and is insensitive to the outcomes an action will presently bring. A model-free learner stores cached values for previously performed actions and selects actions with the highest cached value.

Because model-free learning is computationally simpler but less flexible than model-based learning, it may give rise to behavior that has features of habits. For instance, an animal might continue to press a food lever despite being fully sated because the action itself was previously rewarded and thus associated with high reward value (Dickinson & Balleine, 1994; Daw et al., 2011). Given these characteristics, the model-based/model-free distinction has been used in

several recent studies to probe the role of habits in a range of learning contexts in humans. For instance, individuals who engage in greater model-based learning show less habitual persistence in a devaluation task—a classic marker of habits (Gillan, Otto, Phelps, & Daw, 2015). Yet, to date, this approach has not been applied to questions on the formation of social impressions through direct social interactions with other people.

Model-free learning in social cognition

How might a model-based/model-free account relate to social impressions? When other people us with provide material feedback (like a gift) or social feedback (like a smile or a compliment), we experience this feedback as rewarding; as a result, this feedback can reinforce our social choices and draw us back to the same partners again in the future (Jones et al., 2011; Lin, Adolphs, & Rangel, 2011; Hackel et al., 2015; Lindström, Selbing, Molapour, & Olsson, 2014; Lindström & Tobler, 2018). If people learn from this feedback in a model-free manner, specifically, they might return to interaction partners associated with high reward regardless of whether those partners will provide desirable outcomes in the current context. This pattern would resemble a traditional definition of habit.

Some existing work hints at the possibility that reward feedback gives rise to social preferences that persist in a habit-like manner. In research by Hackel et al. (2015), participants played an economic game in which they chose partners who could share money; partners varied in the average *amount* they shared (indicating reward value) and average *proportion* they shared (indicating generosity). During initial learning, it was economically advantageous for participants to prefer individuals who provided large rewards, regardless of their generosity. However, when participants were later asked to choose one of these partners to work with in a non-economic puzzle-solving task—a context where generosity, but not previous reward value, is advantageous—participants' choices were still influenced by partners' past reward value in addition to their generosity. This persistent influence of past reward—even when reward value no longer informed

desired outcomes—suggests that participants may have developed model-free reward associations that guided subsequent social preferences. Nevertheless, past work has not directly tested this possibility by dissociating model-based and model-free learning in social interaction.

Study overview

The present research was designed to provide initial evidence for model-free learning in social impression formation. To this end, we administered a sequential choice task commonly used to dissociate model-based and model-free learning (Kool, Cushman, & Gershman, 2016; Kool, Gershman, & Cushman, 2017; see also Daw et al., 2011), adapted to examine social partner choice and attitudes. On each round, participants chose financial advisors who had supposedly invested in one of two stocks; participants then received a payout from that advisor's stock. We examined the extent to which participants chose advisors based on model-based and model-free reinforcement, and further examined whether these forms of learning predicted participants' subjective attitude toward each advisor.

Method

Participants

Sixty-nine participants were recruited via Amazon Mechanical Turk (AMT), in exchange for $3.50 for study completion, plus a monetary bonus based on their task performance. A sample size of 65 participants was chosen a priori; an additional four participants completed the task due to an error in which an extra set of slots was posted. Data collection was completed before analysis. Participants were eligible if there were located in the United States, completed at least one prior AMT study, and had approval rates of at least 95%. Informed consent was obtained from all participants in accordance with the guidelines of the New York University Committee on Activities Involving Human Subjects. We excluded data from participants who did not respond in time to either the first or second stage of a trial on more than 20% of trials (Kool et al., 2017). This rule excluded data from four participants, leaving data from 65 participants in analyses.

Procedure

Participation took place via Psiturk, an online platform for cognitive tasks (Gureckis et al., 2016). After providing consent, participants read a self-guided description of the study, which included practice trials, and completed the main experimental task. Next, participants completed self-reported evaluation items and a demographics questionnaire. Lastly, participants were informed of their compensation for participating and then completed a debriefing procedure that included a suspicion probe and an explanation of study goals. All data exclusions, all manipulations, and all measures included in this research are fully reported in this article.

Two-step task. We adapted a sequential learning task (Kool et al., 2016, 2017) designed to dissociate model-free and model-based learning (Figure 1). In our adaptation, participants were told they would learn about choices made by four AMT workers who previously participated in a financial decision-making study. According to this cover story, these previous workers were assigned the role of "Financial Advisor," in which they chose (only) one of two stocks ("Axiom" and "Zephyr") to invest in for the duration of the study. These Advisors then earned money based on the performance of their chosen stock, which fluctuated throughout the study and could change from one round of "dividends" to the next.

Next, participants were assigned to the role of the "Client," in which they would make a series of decisions about which Advisor to hire. Participants learned they would earn points based on the performance of the stock chosen by their hired Advisor on each round. Participants were explicitly told that the performance of the stocks would change over time ("a stock that was bad at the beginning of the game might start performing well, and a stock that initially pays well might perform poorly later on"), and that they should try to hire Advisors with the better performing stock at that particular moment. Moreover, participants were informed that they would receive a monetary bonus for their performance in the task, with better performance (in terms of points earned) equating to a larger bonus.

On each trial, participants began in one of two randomly chosen first-stage states. In these states, participants were presented one of two pairs of Advisors, represented by distinct cartoon avatars (Figure 1). Avatars were randomly assigned to different roles across participants (i.e., which stock they were linked with) and were equally likely to appear on the left or right side of the screen. Participants chose one of the two Advisors via button response and then transitioned deterministically to one of the two stocks, which comprised the second-stage states. That is, participants could reach either of the two stocks from each of the first-stage states; one Advisor in each pair always invested in the Axiom stock and the other Advisor in the given pair always invested in the Zephyr stock.

When they reached the second-stage state, participants were instructed to press the spacebar to reveal the performance of the stock in which the chosen Advisor invested. If participants did not respond in time to either the first- or second-stage states, no reward was provided and participants moved on to the subsequent trial. The number of points obtained for each stock fluctuated slowly and stochastically over the course of the task, varying according to a Gaussian random walk (SD = 2) with reflecting bounds at 0 and +9 points. The drifting nature of the reward feedback encouraged continuous learning throughout the task.

Importantly, the two first-stage states were equivalent in terms of the stocks they could lead to: within each pair of advisors, one Advisor always invested in the Axiom stock, whereas the other Advisor always invested in the Zephyr stock. This design allows for the separation of model-free and model-based control. Given that both stocks can be reached from each pair of Advisors, the stock reached from one set of advisors can be used by a model-based learner to update preferences regardless of which set of advisors is encountered on the next trial. In other words, if an Advisor in one pair invested in the Axiom stock, which paid out a large number of points on that trial, a model-based learner should subsequently be more likely to choose the Advisor in the other pair that also invests in the Axiom stock—a model-based learner can generalize across equivalent first-stage

choice options due to its exploitation of the overarching task structure. Conversely, model-free learners would not generalize across equivalent first-stage choice options, as they simply rely on directly-experienced action-outcome associations—the outcomes experienced following a choice in one pair of advisors should not affect preferences for the advisors in the second pair, and vice-versa.

Participants were trained extensively on the deterministic transitions (i.e., which financial advisor in a given pairing invested in which of the two stocks) prior to completing the experimental trials, such that 80% accuracy across 15 consecutive trials was required to advance to the main task. Afterward, participants completed 150 trials of the main task, split evenly between the two first-stage states. The response deadline in both stages was 1500ms and feedback was presented for 1000ms.

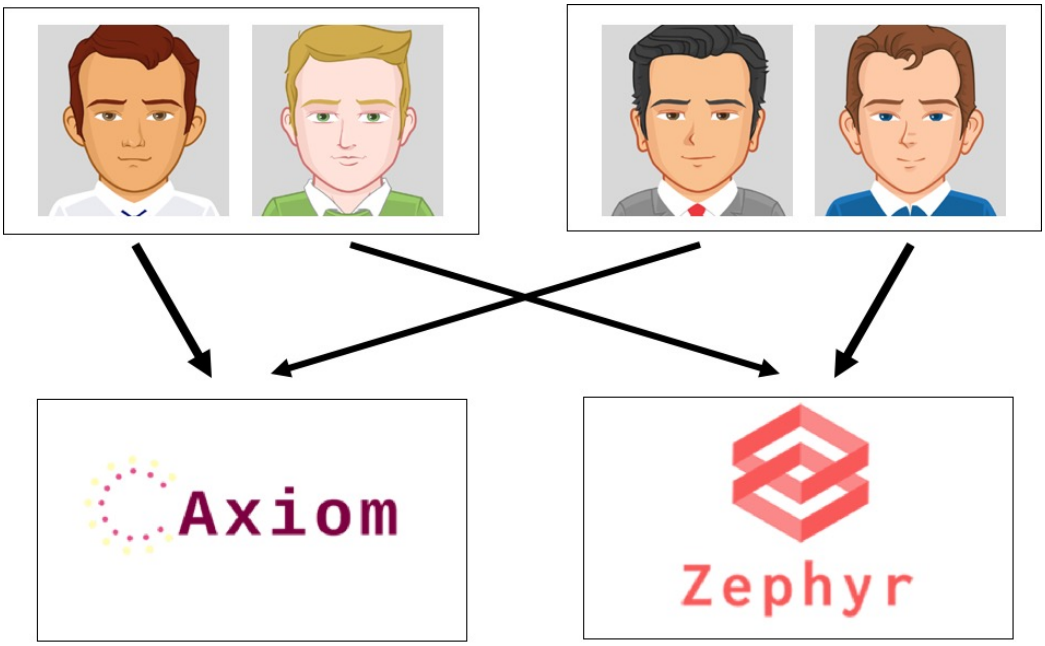

Figure 1. Schematic of task design. In the first stage of each round, participants saw one of two sets of advisors and chose an advisor for that round. Participants then viewed the stock that advisor had chosen; after making a button press, participants saw feedback indicating the payout provided by the stock, ranging from zero to nine points. Within each pair of advisors, one advisor always led to the "Axiom" stock and the other always led to the "Zephyr" stock. This feature of the task rendered the two sets of advisors equivalent, such that a model-based learner could apply experiences with one set of advisors to choices involving the other set of advisors.

Post-task evaluations. Following the two-step task, participants responded to a series of self-report items, which pertained to participants' evaluations (or "liking") of the different Advisors encountered during the two-step task. Participants were presented with the avatar of each financial advisor, one at a time, and rated how much they liked the advisor using a seven-point scale (from 1 = "Do not like them at all" to 7 = "Like them a lot").

Participants were also asked to estimate how valuable, on average, each of the two stocks were over the course of the learning task. These data are beyond the scope of this paper.

Computational model

In order to determine the degree to which participants employed model-based and model-free learning, we fit data from the learning phase to a computational model of reinforcement learning used in previous work (Kool et al., 2017). Doing so allowed us to estimate latent variables related to social learning for each subject (Hackel & Amodio, 2018), which we then used as input in our analyses.

The model contains a hybrid of model-free learning and model-based learning for selecting advisors (see Supplemental Materials for additional details and Table S1 for parameter fits). The model-free system stores values for advisors at the first stage and for stocks at the second stages based on prior reward feedback. The model-based system computes the value of selecting each advisor at the time of choice, combining knowledge about how advisors lead to stocks with the expected payoff of each stock (acquired through model-free learning at the second stage). A model-based learner thus prospectively plans towards a goal: he or she selects an advisor based on the stock the advisor will lead to, in light of the reward expected from each stock. In contrast, a model-free learner values advisors based on the rewards those advisors have led to in the past.

Critically, the model includes a weighting parameter (w) that indicates the relative influence of model-based and model-free learning in choice, ranging between 0 (purely model-free) and 1

(purely model-based). This parameter can serve as an individual difference measure of the extent to which a participant engaged in model-based or model-free learning. We fit this model for each participant using Maximum a Posteriori estimation, with empirical priors used in previous work (Kool et al., 2017). Doing so allowed us to estimate each participant's *w* parameter (median = .92), indicating the extent to which they relied on model-based versus model-free learning. We used this parameter in subsequent analyses examining individual differences in the use of these learning strategies.

Results

Model-free and model-based social learning

Did participants engage in either model-based or model-free social learning? To answer this question, we examined choices in the learning phase, drawing on the following logic of the task. As noted above, the two sets of advisors in the task are equivalent, such that one advisor from each set leads to each stock. As a result, a model-based learner would generalize experiences with one set of advisors to the other set. For instance, imagine a participant who sees the first pair of advisors, picks the advisor that leads to the "Axiom" stock, and receives a large reward. On the next round, a model-based learner would try to return to the "Axiom" stock regardless of whether they see the same pair of advisors or a different pair of advisors. In contrast, a model-free learner updates values for individual advisors and chooses advisors based on these values. A model-free learner would therefore repeat their choice on the next trial if presented with the same advisors but would do so to a lesser extent if presented with different advisors. That is, the model-free learner would fail to generalize across sets of advisors.

Drawing on this task logic, we fit learning phase data to a lagged regression model predicting, on a trial-by-trial basis, whether or not participants repeated their most recent choice of Stage 2 stocks (1 = stay, 0 = switch). Following Kool et al. (2016), predictors included the reward

earned on the previous trial (standardized, within-subject, to z-scores), whether or not the previous trial started with the same set of advisors (1 = same, -1 = different), and the interaction of these two predictors. A main effect of reward indicates model-based learning: people return to a high-paying stock, regardless of whether they see the same or different advisors on the next trial to get to that stock (simulated data shown in Fig 2a). An interaction of reward and start state indicates model-free learning: people try to return to a high-paying stock, but particularly do so when presented with the same set of advisors and can therefore repeat the advisor choice that led to the large reward (Fig 2b). Models were fit using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2016). Random variances were allowed for the intercept and all slopes. (See Table S2 for all coefficients.)

This analysis produced a main effect of reward, $b = 1.47$, $SE = .07$, $z = 19.80$, $p < .001$, consistent with model-based learning: overall, participants returned to second-stage stocks after receiving large rewards. However, the analysis also produced a Reward x Start State interaction, $b = .22$, $SE = .03$, $z = 6.45$, $p < .001$, indicating the presence of model-free learning: participants were more likely to return to a high-paying stock when starting with the same advisors at the first stage. Our data thus supported the hypothesis that both model-based and model-free reinforcement learning contributed to social choice (Fig. 2c), consistent with a role for habitual learning in social preferences in addition to goal-directed learning.

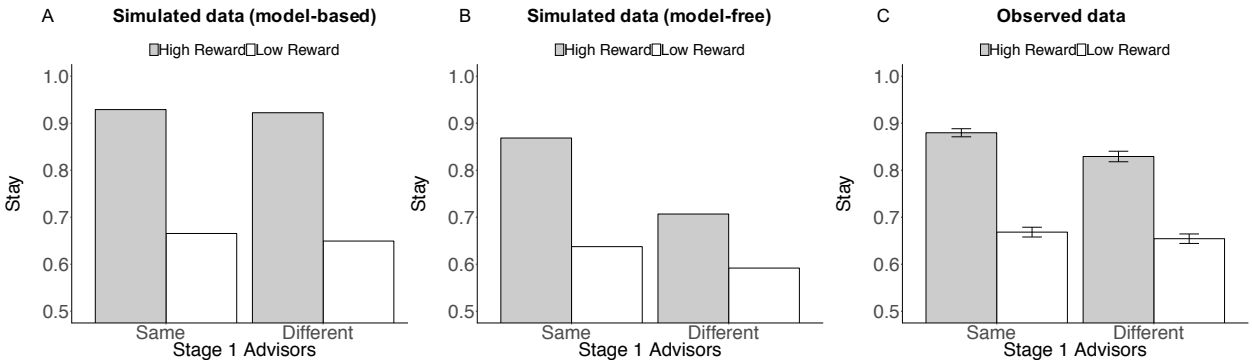

Figure 2. Behavioral predictions and data. Plots depict the probability of staying with the same second-stage stock as on the previous trial, based on whether the set of advisors encountered at the first stage was the same as or different from that of the previous trial, and whether the previous trial delivered a high or low reward. (A) Simulated model-based predictions. (B) Simulated model-free predictions. Panels A and B produced from model simulations (see Supplemental Materials) with weighting parameter *w* specifying fully model-based (w = 1) and fully model-free learning (w = 0), respectively. (C) Observed data indicates the presence of both model-based and model-free learning. Error bars reflect standard error of the mean, adjusted for within-subjects comparisons (Morey, 2008).

Post-task evaluations

If reinforcement learning also gives rise to attitudes, participants might like advisors who can provide reward in the future (model-based value) and advisors associated with past reward (model-free value). To test how learning impacts attitudes, we examined self-reported liking of each advisor completed after the learning task. Using each subject's individual parameter fits in the computational model, we estimated the final model-based and model-free value associated with each advisor for each subject at the end of learning, given the unique series of stimuli and outcomes viewed by each participant. We then regressed liking ratings simultaneously on each type of value.

Notably, model-based values were identical for advisors who led to the same stock. That is, if the Axiom stock would be expected to deliver 6 points on average at the end of the task, then each advisor who leads to the Axiom stock would have a model-based value of 6 points. If social evaluations reflect model-based learning, participants would therefore equally like two advisors who led to the Axiom stock. In contrast, model-free values reflect the unique reward history associated with a particular advisor; even for two advisors who led to the Axiom stock, participants

might have experienced different reward outcomes with each advisor. If social evaluations reflect model-free learning, people would therefore also prefer advisors who provided greater rewards. Finally, this tendency should depend on individual differences in learning, as reflected in the *w* parameter: individuals who engage in greater model-free learning should especially like advisors associated with high model-free value.

To test these hypotheses, we fit a mixed-effects linear regression predicting post-task liking ratings (Table S3). Predictors included each participant's final model-free values and model-based values towards each advisor (estimated from the computational model), each participant's *w* parameter, and the interaction of *w* with each type of value. Each predictor was standardized to z-scores (within-subject for the value regressors and between-subject for the *w* parameter). As a result, main effects of value regressors are interpretable relative to the mean level of the *w* parameter (w = .83). We included random variances for the intercept and each predictor. The models were fit using the lme4 package and lmerTest packages (Bates, Maechler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2016) in R (R Core Team).

This analysis yielded a main effect of model-based values, b = .30, SE = .14, $t(71.46) = 2.17$, p = .03, and a marginally significant main effect of model-free values, b = .16, SE = .09, $t(162.97)$ = 1.82, p = .07. In other words, at mean levels of the *w* parameter, people liked advisors who could lead them to more rewarding stocks and also liked advisors who were uniquely associated with greater reward in the past.

We further examined whether the effects of model-based and model-free learning on reported attitude varied by participants' individual learning tendencies, as indexed by the *w* parameter. We found that the *w* parameter, which represents this individual difference variable, interacted with model-free values, b = -.24, SE = .08, $t(148.01) = -2.97$, p = .004. Participants who exhibited relatively greater model-free learning also expressed greater liking of partners who had provided more reward. Simple effects analysis supported this interpretation: for model-free

learners (centered at the 25th percentile of the *w* parameter, or *w* = .70), model-free values were strongly predictive of attitudes towards advisors, *b* = .31, *SE* = .10, *t*(155.32), = 3.11, *p* = .002, revealing a novel effect of model-free learning on social evaluation. By contrast, for model-based learners (centered at the 75th percentile of the *w* parameter, or *w* = 1), model-free values were not associated with evaluations, *b* = -.03, *SE* = .11, *t*(162.01)= -.31, *p* = .76. Thus, participants who exhibited model-free learning also liked advisors associated with greater model-free value.

Together, these results identify two ways in which reinforcement learning influences social attitudes: people like others who are equivalently capable of providing large rewards in the future and people like others who have uniquely provided large rewards in the past. Moreover, the influence of past (model-free) reward history depends on individual differences in learning: individuals who weight model-free rewards more strongly during learning also have a stronger preference for advisors associated with past rewards.

Discussion

Does habit play role a social impressions? Our findings demonstrate that, indeed, people form impressions through reward-based reinforcement processes that include model-free learning—a form of learning thought to contribute to habitual behavior. In the sequential learning task used here, participants chose financial advisors based on both model-based and model-free learning. That is, participants chose advisors based on who could lead them to desirable stocks in the future (model-based) as well as who was uniquely associated with high rewards in prior interactions (model-free). This pattern of model-free learning, in particular, suggests a habit-like component of learning and behavior in the context of social impression formation.

Furthermore, learning related to participants' explicit social evaluations. Across participants, both model-based and model-free learning predicted self-reported attitudes toward advisors. Moreover, participants varied in their reliance on model-based vs. model-free processing during the learning task, and this individual difference in learning related to differences in

evaluation: participants who exhibited greater model-free learning during the investment task showed an effect of model-free learning on self-reported attitudes. Thus, these findings dissociate two routes by which reinforcement learning contributes to attitudes, and they highlight the importance of considering individual differences in learning strategies during social interactions to understand the effects of rewards on social attitudes and decisions.

Model-based and model-free social cognition

Our central finding—of model-free learning in social impression formation—offers novel theoretical implications for social cognition, learning, and attitudes. First, our findings highlight a role for reward-based reinforcement learning in social interactions. Previous impression formation research demonstrates that people learn about the traits of others in order to predict how others will behave (Heider, 1958). For instance, by observing financial advisors, people can form impressions of an advisor's competence and predict that advisor's future performance (Boorman et al., 2013; Leong & Zaki, 2018). Our results introduce a complementary mode of social learning based on reward: people also learn who to choose and who to like through instrumental learning, such as directly choosing an advisor and experiencing rewards as a result.

The observation of model-free social learning, in particular, supports the proposed role of habit in social cognition. In model-free learning, people repeat previously rewarded choices in a relatively inflexible manner—the hallmark of a habit. Habits may therefore influence social behavior: because habits reflect routinized responses that operate most adaptively in invariable environments, they may fill in the gaps between goal-directed responses to facilitate social behavior. In some cases, habits may have harmful effects; for example, people may persist in interacting with social partners with whom they had positive past experiences, even when other partners might be equally or more relevant to one's current goals. In other cases, habits may be beneficial, leading an individual to approach a previously-rewarding person while distracted by their pursuit of an unrelated goal--perhaps eliciting help, if needed, or simply avoiding a social faux

pas. In both cases, their effects may be subtle, relative to goal-directed responses, yet still crucial to adaptive social function.

Although model-based and model-free learning offer different benefits and costs, their concerted function may promote successful social interactions. Social life offers a wealth of information about other people—their traits, preferences, and emotions—which lets us know whom to interact with and how to interact with them. Through experience, we learn which members of our social networks to turn to for empathy as opposed to fun (Morelli, Ong, Makati, Jackson, & Zaki, 2017) and which verbal or facial cues predict different emotions for close others (Zaki, Kallman, Wimmer, Ochsner, & Shohamy, 2016). Model-free learning offers a computationally simple way to learn how to act around others given this wealth of information, requiring little deliberation or careful thought (Otto, Gershman, Markman, & Daw, 2013). Yet, at the same time, model-free learning is relatively inflexible, leaving people unable to adapt as contingencies change or to plan ahead in novel settings. By comparison, model-based learning requires greater effort but allows people to adapt to new contingencies and make novel plans—for instance, choosing a gift for another person for the first time given knowledge about their preferences. Both types of learning are functional, with tradeoffs that depend on the particulars of a situation, and thus an important goal of future research will be to explore how these tradeoffs are managed and prioritized across situations.

Finally, and more broadly, this work sheds light on how multiple forms of learning and memory can contribute to social cognition. Based on research in cognitive neuroscience (Squire, 2004; Henke, 2010), Amodio (2019; Amodio & Ratner, 2011) theorized that social cognition comprises multiple distinct and interactive learning and memory systems, including habits. Although classic work in social psychology has focused primarily on the roles of conceptual associations and Pavlovian forms of learning, research has just recently begun to probe the role of reward-based forms of learning in social cognition (Hackel et al., 2015; Lindström & Tobler, 2018).

To date, these studies have not distinguished between types of computations that may underlie instrumental learning from rewards. Here, by using a two-step learning task to examine social learning, we were able to dissociate model-based and model-free forms of reward learning and, in doing so, provide new evidence for the role of multiple learning systems, functioning in concert, in social cognition.

Limitations

The goal of this research was to examine learning processes that give rise to habitual behavior. However, there remain open questions about the extent to which model-free learning, as assessed in sequential decision-making (i.e., two-step) tasks, corresponds to traditional definitions of habit. First, questions have been raised as to whether additional strategies may contribute to observed effects of model-free learning in sequential decision tasks (Da Silva, Yao, & Hare, 2018; Dezfouli & Balleine, 2012; but see Morris & Cushman, 2019). Second, it is a matter of debate whether—and to what extent—model-free learning maps on to traditional definitions of habitual control (Miller, Shenhav, & Ludvig, 2019; see also Sjoerds et al., 2016; Gillan et al., 2015). Miller and colleagues (2019) argue that traditional conceptualizations of habits reflect stimulus-response associations devoid of expected value representations (i.e., are *value-free*), whereas model-free algorithms still depend on the expected value representations associated with a learner's available actions (i.e., are *value-based*). In this view, habits form directly through action repetition within a given context, regardless of reward outcomes.

It is possible that both model-free RL and action repetition contribute to behaviors commonly considered habitual (Pauli, Cockburn, Pool, Pérez, & O'Doherty, 2018). These processes might align with a theoretical distinction between "direct" cuing of habits, in which responses are directly associated with context cues, and "motivated" cuing of habits, in which responses depend on the motivation linked to a behavior through past rewards (Wood & Neal, 2007). To complement and extend our findings, future work could consider these varied approaches.

New questions about habits in social behavior

Our use of the two-step task to probe the role of habits in social cognition raises several new questions regarding other aspects of habits in social life. For instance, a classic marker of a habit is its persistence even when it no longer fulfills a valued goal (Wood & Rünger, 2016). Past work suggests that reward feedback in social interaction can have such a persistent impact (Hackel et al., 2015). Future work should consider tasks traditionally employed to test for this kind of habitual persistence, such as the slips-of-action paradigm (e.g., de Wit et al., 2012; Gillan et al., 2011) or outcome devaluation/revaluation procedures (e.g., de Wit, Corlett, Aitken, Dickinson, & Fletcher, 2009; Tricomi, Balleine, & O'Doherty, 2009; Valentin, Dickinson, & O'Doherty, 2007).

In real-world settings, social habit strength can be measured with self-report assessments and reaction times that reveal the accessibility of habitual responses following a context cue (Wood & Rünger, 2016). These measures can allow researchers to probe the existence and structure of social habits in everyday relationships. For instance, do people form habits to interact with specific partners in specific contexts? Or do they form habits to approach or avoid social interaction in general? Are there benefits to forming such social habits? Answering these questions promises to illuminate the structure of people's social lives, much as advances in habit research sheds light on how habits can promote healthy eating, exercising, or studying (Galla & Duckworth, 2015; Lin, Wood, & Monterosso, 2016).

Finally, the implications of our findings may extend to other areas of research within social psychology, such as intergroup relations. The concept of habit has been invoked in prior theories of social attitudes, such as to describe the phenomenon of implicit prejudice and the difficulty people have in ridding themselves of it (e.g., "breaking the prejudice habit," Devine, 1989; Devine, Forscher, Austin, & Cox, 2012). However, this usage has been largely colloquial or metaphorical, as previous research has not used methods capable of assessing habit-like patterns of preference and choice. Our findings suggest that social experiences may indeed give rise to a form of habit, but

these are rooted more directly in action tendencies than in conceptual processes such as stereotypes.

Nevertheless, if some aspects of prejudice are truly habit-like, then they may be extraordinarily difficult to control or eradicate. As such, interventions involving the replacement of a biased thought or action with an egalitarian response (Devine, 1989) or changes in the situational affordances for bias expression (Amodio & Swencionis, 2018) should be more effective than methods for unlearning bias (Lai et al., 2014). Furthermore, an intervention aimed at "unlearning" of a habit-like response would require action-based interventions, in contrast to conventional interventions aimed at modifying a person's beliefs and values. As our conceptualization of habits in social cognition develops, it may begin to to elucidate psychological processes in other domains as well.

Conclusion

Habits are integral to everyday human behavior, and they may also support our social behaviors. Our findings represent an initial demonstration that habit-like learning processes are also involved in the formation of social preferences and attitudes. These findings expand our understanding of how learning and memory systems support social cognition and provide a foundation for new research on the role of habit in social learning.

References

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

Amodio, D. M., & Berg, J. J. (2018). Toward a multiple memory systems model of attitudes and social cognition. *Psychological Inquiry, 29,* 14-19.

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

Amodio, D. M., & Swencionis, J. K. (2018). Proactive control of implicit bias: A theoretical model and implications for behavior change. *Journal of Personality and Social Psychology*, *115*(2), 255.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*, 258-290.

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

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, 126, 925-945.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Boorman, E. D., O'Doherty, J. P., Adolphs, R., & Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron*, *80*(6), 1558-1571.

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

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.

Da Silva, C. F., Yao, Y. W., & Hare, T. A. (2018). Can model-free reinforcement learning operate over information stored in working-memory?. *BioRxiv*, 107698.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*, 1204-1215.

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition*, *6*(1), 87-107.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267-1278.

de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *Journal of Neuroscience*, *29*(36), 11330-11338.

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience*, *32*(35), 12066-12075.

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience*, *35*(7), 1036-1051.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1-18.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*, 767-772.

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.

Fitzsimons, G. M., & Anderson, J. (2013). Interpersonal Cognition: Seeking, Understanding, and Maintaining Relationships. In D. Carlston (Ed.) *Handbook of social cognition* (pp. 590-615). New York: Oxford University Press.

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology* 109, no. 3 (2015): 508.

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry*, 168(7), 718-726.

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523-536.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., ... & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829-842.

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

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233.

Heider, F. (1958). The psychology of interpersonal relations. New York, NY: Wiley.

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

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*(3), 390.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York, NY: Academic Press.

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., ... & Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *The Journal of Neuroscience*, *31*, 13039–13045.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off?. *PLoS Computational Biology*, *12*, e1005090.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, *28*, 1321-1333.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-32.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., ... & Frazier, R. S. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765.

Leong, Y. C., & Zaki, J. (2018). Unrealistic optimism in advice taking: A computational account. *Journal of Experimental Psychology: General*, *147*(2), 170.

Lin, A., Adolphs, R., & Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*, 274-281.

Lin, P. Y., Wood, W., & Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite*, *103*, 432-440.

Lindström, B., Selbing, I., Molapour, T., & Olsson, A. (2014). Racial bias shapes social reinforcement learning. *Psychological Science*, *25*(3), 711-719.

Lindström, B., & Tobler, P. N. (2018). Incidental ostracism emerges from simple learning mechanisms. *Nature Human Behaviour*, *2*, 405-414.

Miller, K. J., Ludvig, E., Pezzulo, G., & Shenhav, A. (2018). Re-aligning models of habitual and goal-directed decision-making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-directed decision making: Computations and neural circuits*. Amsterdam, the Netherlands: Elsevier.

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review*, *126*, 292-311.

Morelli, S. A., Ong, D. C., Makati, R., Jackson, M. O., & Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, *114*(37), 9843-9847.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.

Morris, A., & Cushman, F. A. (2019, April 5). Model-free RL or action sequences?. https://doi.org/10.31234/osf.io/k67tm

Moskowitz, G. B. (1993). Person organization with a memory set: are spontaneous trait inferences personality characterizations or behaviour labels?. *European Journal of Personality*, *7*, 195-208.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421-433.

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, *24*(5), 751-761.

Pauli, W. M., Cockburn, J., Pool, E. R., Pérez, O. D., & O'Doherty, J. P. (2018). Computational approaches to habits in a model-free world. *Current Opinion in Behavioral Sciences*, *20*, 104-109.

R Core Team. (2016). R: A Language and Environment for Statistical Computing - Version 3.3.1.

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

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.

Sjoerds, Z., Dietrich, A., Deserno, L., De Wit, S., Villringer, A., Heinze, H. J., ... & Horstmann, A. (2016). Slips of action and sequential decisions: a cross-validation study of tasks assessing habitual and goal-directed action control. *Frontiers in Behavioral Neuroscience*, *10*, 234.

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

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. Cambridge, MA: MIT Press.

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personality and Social Psychology Review*, *18*(2), 187-209.

Thorndike, E. (1911). *Animal intelligence*. New York, NY: Hafner.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in cognitive sciences*, *22*(3), 201-212.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, *29*, 2225-2232.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). New York, NY: Guilford Press.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*, 490-501.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience*, *27*(15), 4019-4026

Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*(6), 919

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*, 237-252.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*(4), 843-863.

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289-314.

Wyer, R. S., Jr., & Carlston, D. E. (1979). *Social cognition, inference, and attribution*. Hillsdale, NJ: Erlbaum Publishers.

Wyer, R. S., Jr., & Srull, T. K. (1989). Memory and cognition in its social context. Hillsdale, NJ: Erlbaum.

Zaki, J., Kallman, S., Wimmer, G. E., Ochsner, K., & Shohamy, D. (2016). Social cognition as reinforcement learning: Feedback modulates emotion inference. *Journal of Cognitive Neuroscience*, *28*(9), 1270-1282.