Journal of Experimental Social Psychology 88 (2020) 103948

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

Reinforcement learning in social interaction: The distinguishing role of trait inference

Leor M. Hackel[a,*], Peter Mende-Siedlecki[b], David M. Amodio[c,d]

[a] University of Southern California, United States of America
[b] University of Delaware, United States of America
[c] New York University, United States of America
[d] University of Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Social cognition
Trait
Learning
Reward

ABSTRACT

People learn about the world by making choices and experiencing feedback—a process characterized by models of reinforcement learning in which people learn to associate their actions with rewarding outcomes. Although reinforcement models provide compelling accounts of feedback-based learning in nonsocial contexts, social interactions typically involve inferences of others' trait characteristics, which may be independent of their reward value. As a result, people may learn differently about humans and nonhumans through reinforcement. In two experiments (and a pilot study), participants interacted with human partners or slot machines that shared money. Computational modeling of behavior revealed different patterns of learning for humans and non-humans: participants relied more on feedback indicating trait generosity (relative to monetary reward) when learning about humans but relied more on monetary reward (relative to generosity) when learning about slots. Furthermore, this pattern of learning had implications for attitudes: whereas participants preferred generous humans, they preferred rewarding slot machines, relative to their respective counterparts. These findings reveal a distinct role for reinforcement learning in social cognition, showing that humans preferentially form abstract trait inferences about other people through feedback in addition to reward associations.

1. Introduction

We often learn about other people through trial and error: we engage with them and learn from their responses. These interactions resemble reinforcement learning, in which people and animals learn by making choices and experiencing feedback (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Sutton & Barto, 1998). For instance, rats repeat lever presses that lead to food rewards (Balleine & Dickinson, 1998) and humans return to slot machines that pay out cash (Daw, O'Doherty, Dayan, & Seymour, 2006). But how well do such models characterize social interaction? While existing models of reinforcement explain how people learn from rewards in non-social contexts, human social interactions often involve more abstract inferences about others' traits—inferences that guide social behavior beyond reward associations. Rather than returning solely to others who provide concrete rewards, people may also return to those who display valuable traits. In this research, we examined how social contexts transform the dynamics of reinforcement learning, asking whether people are more likely to learn from traits, over and above rewards, in social compared with nonsocial interactions.

1.1. Traits as a source of value

Perceivers frequently attribute stable traits to others, which can be used to predict others' future behavior and thus estimate the value of a social interaction (Hamilton, Katz, & Leirer, 1980; Hastie, 1980; Heider, 1958; Rim, Uleman, & Trope, 2009; Srull & Wyer, 1989; Winter & Uleman, 1984). For example, a colleague viewed as "generous" can be expected to treat us kindly across a range of situations. Although people with desirable traits may also typically provide rewarding outcomes—for example, when a generous colleague treats us to lunch—traits and rewards can diverge, as when a generous colleague has few material resources to share.

Research shows that in social interactions, people learn to associate

* Corresponding author at: Department of Psychology, University of Southern California, 3620 South McClintock Ave, Los Angeles, CA 90089, United States of America.

E-mail address: lhackel@usc.edu (L.M. Hackel).

https://doi.org/10.1016/j.jesp.2019.103948
Received 23 April 2019; Received in revised form 17 December 2019; Accepted 29 December 2019

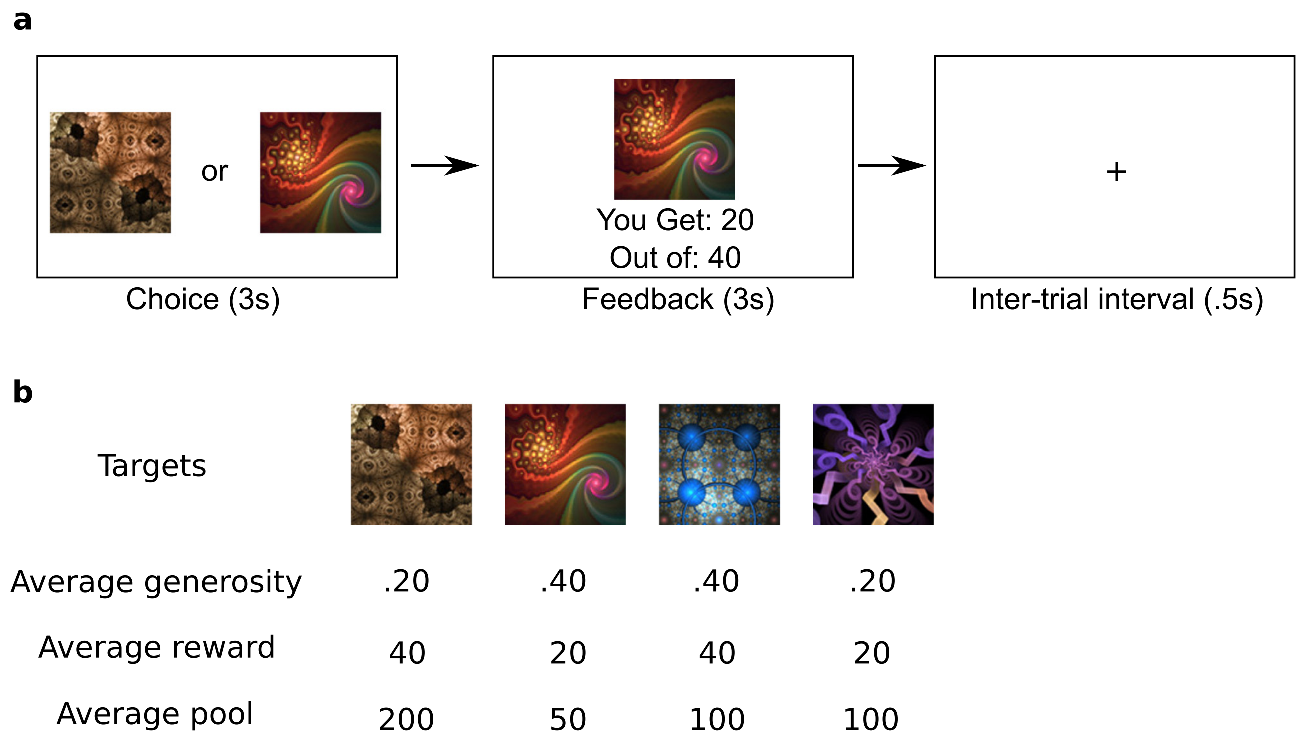

Fig. 1. Schematic of learning task. (a) On each round, participants chose to interact with one of two target characters represented by fractal images, described as other humans or slot machines. Feedback indicated the reward obtained from the target and the available point pool, which in turn indicated the target's generosity (i.e., proportion shared). (b) Targets varied independently in the average reward and generosity with which they were associated.

others with both rewards and traits. In a study by Hackel, Doll, and Amodio (2015), participants learned about humans who shared money and slot machines that paid money (in the form of points later exchanged for cash). Some humans and slots offered larger *amounts* of money than others (offering high reward value) and, independently, some offered larger *proportions* of available money than others (indicating trait-level generosity). For instance, one target shared 20 points out of 100, on average, whereas another shared 10 points out of 25; the former offered a larger reward whereas the latter was more generous. Previous models of reinforcement would predict that rewards alone should reinforce behavior: people should return to partners who provide large amounts of money. Yet, participants in Hackel et al. (2015) learned to select partners based on both reward and trait feedback, as indicated by their behavior and by neural activity in ventral striatum, a region strongly linked to reinforcement (Garrison, Erdeniz, & Done, 2013). Moreover, people relied more strongly on generosity feedback than reward feedback when choosing partners for future interactions. Hence, this research demonstrated that both reward and trait information may reinforce choice.

In the present research, we asked whether this capacity to infer traits through reinforcement, in addition to rewards, is enhanced in social contexts, in which case it may reveal a functional role of social cognition in reinforcement learning. Although people may attribute abstract, stable characteristics to both humans and objects (Baetens, Ma, Steen, & Van Overwalle, 2013; Baetens, Ma, & Van Overwalle, 2017; Heider, 1958; Riva, Sacchi, & Brambilla, 2015; Spunt & Adolphs, 2015; Waytz, Heafner, & Epley, 2014) and learn from rewards involving both (Lin, Adolphs, & Rangel, 2011), traits provide a richer representation of conceptual information, beyond reward value, which may help people flexibly choose partners relevant to one's current goals (Amodio, 2019; Hackel, Mende-Siedlecki, Loken, & Amodio, 2019).

Although prior research by Hackel et al. (2015) assessed trait and reward learning from both human and nonhuman agents, it was not optimized for determining whether these two forms of reinforcement learning are differentially engaged in social and nonsocial contexts. For example, in the task used by Hackel et al. (2015), trials alternated between human and slot machine interactions, potentially blurring the social and nonsocial nature of interactions or leading participants to anthropomorphize slots (Hsu & Jenkins, 2015). Furthermore, human and slot machine stimuli were not equated on factors that would have permitted a strong direct test of social context (e.g., visual appearance). Indeed, no significant differences between social and non-social target interactions were observed in participants' neural activity or in computational modeling of learning, and small differences observed in participants' choice behavior may have been due to extraneous factors (e.g., better memory for face stimuli than slot stimuli; Gauthier & Nelson, 2001). Hence, the results of Hackel et al. (2015) did not directly inform whether different forms of learning are engaged to support social as opposed to nonsocial cognition. In fact, the lack of clear differences observed by Hackel et al. (2015) raised the question of whether trait-level reinforcement related to social cognition at all; instead, “trait-learning” might have reflected a domain-general process of learning from relative (versus absolute) rewards, wherein people simply contrast rewards to contextual baselines (Holroyd, Larsen, & Cohen, 2004; Kahneman & Tversky, 1979; Palminteri, Khamassi, Joffily, & Coricelli, 2015; Rigoli, Rutledge, Dayan, & Dolan, 2016). Due to these ambiguities, the question of whether reinforcement learning differs across social versus nonsocial interactions remains unaddressed.

The present work was designed explicitly to test whether trait- and reward-based forms of learning are differentially engaged by social and nonsocial agents. To this end, we examined whether social contexts transform how people learn from feedback, such that people rely more on trait learning, relative to reward learning, in social versus nonsocial interactions. By doing so, we sought to understand how reinforcement learning contributes to interactive social behavior and how social learning differs from non-social learning.

2. Study 1

In Study 1, we tested the hypothesis that people prioritize trait-based learning, compared with reward-based learning, when interacting with other humans, but learn comparatively more from rewards when interacting with nonhumans (e.g., slot machines). We further expected these differences in learning to guide explicit evaluations, leading participants to prefer rewarding slot machines and generous humans relative to their respective counterparts.

2.1. Method

2.1.1. Overview

In a pre-registered experiment, participants completed a version of the Hackel et al. (2015) task (Fig. 1), modified such that (a) participants interacted with only humans or slot machines in between-subjects conditions; (b) identical visual stimuli (fractal images) represented humans and slots; (c) the statistical distributions of reward feedback and generosity feedback were matched so that both signals would, in principle, be equivalently easy to learn; and (d) feedback was described using identical language across human and slot interactions, in contrast to past work and a pilot study (reported in Supplemental materials) in which different language was used to describe feedback from humans ("Shared") and slot machines ("Payout"). This modified task design ensured that any differences in learning would be due only to the social/nonsocial task framing.

2.1.2. Participants

One hundred seventy participants were recruited through Amazon's Mechanical Turk (68 female, 101 male, 1 nonbinary; mean age = 35.69, SD = 10.85) in exchange for payment. Sample size was determined with a non-parametric power analysis based on data from a pilot study (see Supplemental materials for details). Data collection was completed prior to analyses. Participants provided informed consent in accordance with approval from the NYU University Committee on Activities Involving Human Subjects.

Based on past research using reinforcement tasks in online populations (Gillan, Otto, Phelps, & Daw, 2015; Hackel & Zaki, 2018), data were excluded from analysis if a participant responded with mean latencies ± 2 standard deviations from the group mean, did not respond on more than 10% of trials, or pressed the same key on more than 90% of trials. These a priori criteria excluded data from 32 participants, leaving 70 participants in the "human" condition and 68 participants in the "slot" condition. Results remained significant with a more lenient rule that excluded fewer participants (see Supplemental materials).

2.1.3. Procedure

Following random assignment to one of two conditions, participants were told they would learn about other humans or computerized slot machines. In the "human" condition, participants were told they would learn about four previous MTurk workers ("Deciders") who had made many decisions to divide pools of points worth money with a future participant ("Recipient"). Participants were informed they would play in the "Recipient" role, and that points they earned by choosing Deciders who shared would be exchanged for money at the end of the game. Participants were further informed that a different pool of points was made available by the experimenters to each Decider on each round, and the Decider could decide how many points to share. Participants were shown four fractal images, which they were told represented the four previous players.

In the "slot" condition, participants received identical instructions, with one exception: they were told they would be learning about four computerized slot machines, which had determined how many points to pay out from pools created by the experimenters. Participants viewed the same four fractal images as in the human condition, but these were said to represent the four slot machines.

Participants then completed the interaction task, beginning with a learning phase (72 trials). On each trial, participants saw two fractal images representing either human players or slot machines. Each possible pairing of images was viewed 12 times; the left-right location of each image was counterbalanced. Participants were required to select an image within 3 s via button-press. Responses were immediately followed by a 3-second feedback display indicating the number of points received (labeled "you get") and the point pool available to the chosen target for that trial (labeled "out of"). The correspondence between the four fractal images and their feedback patterns (Fig. 1) was randomized across participants to control for any effects of a particular fractal stimulus.

Reward and generosity feedback were generated by using the average values displayed in Fig. 1, plus Gaussian noise with SD = 10 for reward and SD = 0.10 for generosity. These standard deviations were equivalent given the difference in scaling of the two distributions, rendering the distributions equally discriminable. Values followed a censored normal distribution, such that reward value had to be at least 2 points and generosity had to be at least 0.01, to ensure meaningful values. To set point pools displayed during feedback, reward feedback was rounded to the nearest integer and divided by generosity, and the resulting pool value was rounded to the nearest integer.

Participants next completed a test phase, in which they made additional choices without receiving feedback. In this phase, the point pool available for each option appeared above each image before choice, so that participants could integrate these point pools with their learned representations of generosity to make decisions. For instance, a Decider who has 100 points available and previously had shared 40% on average would be expected to share 40 points. Participants completed 72 trials of the test phase, again featuring each possible pair of images. To determine the point pools shown above each image, a random integer from 10 to 100 was generated for one target. This amount was multiplied by one of five ratios, symmetric around 1:1 (0.67, 0.9, 1, 1.11, 1.5), to generate the second target's point pool. We included 24 trials at a 1:1 ratio, because these trials may be particularly informative about preferences rooted solely in prior experience. All other ratios appeared 12 times (twice with each image pair).

Following the main task, participants completed three sets of ratings. First, to test how learning related to explicit evaluations, participants rated how much they liked each human or slot machine, using a 7-point Likert-type scale (1 = not at all, 7 = very much). Next, to assess whether participants explicitly used a particular learning strategy, participants reported the extent to which they thought about each target's personality and the number of points offered by each target, using the same seven-point scale (see Supplemental Results). As an exploratory measure, participants completed a questionnaire assessing individual differences in anthropomorphism (Waytz, Cacioppo, & Epley, 2010; see Supplemental Results). Finally, participants completed exploratory questions probing their acceptance of the cover story (see Supplemental Results). Upon completing these measures, participants were paid a bonus based on the number of points they accrued.

Information regarding our procedure for determining sample size, all data exclusions, all manipulations, and all measures included in this research are fully reported in this article. Materials and de-identified data are available at: https://osf.io/gxdeu/. A pre-registration document is available at: https://aspredicted.org/te25g.pdf.

2.2. Results

2.2.1. Computational model

Were participants more likely to learn trait information than reward information when interacting with humans as opposed to slot machines? To test this key hypothesis, we fit behavioral data from the learning and test phases to a computational model validated in previous work (Hackel et al., 2015; Hackel and Zaki, 2018), which modeled trial-by-trial dynamics of learning in order to estimate how participants weighted each type of feedback during learning (Hackel and Amodio, 2018). The model assumes that people update estimates of both reward value and generosity upon receiving trial-by-trial feedback. During

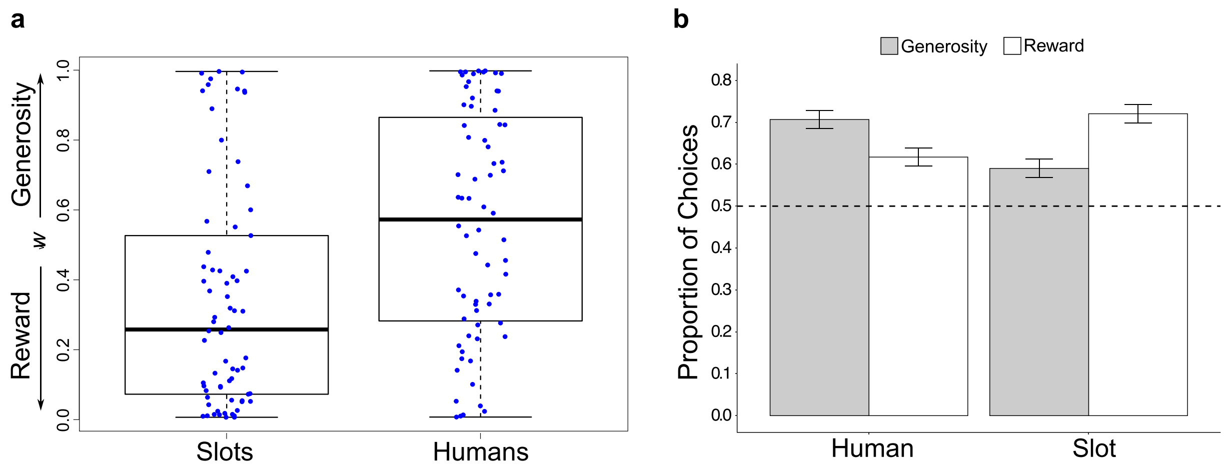

Fig. 2. Study 1 results in the learning and test phases. (a) Degree of reinforcement learning from rewards as opposed to generosity is indicated by a weighting parameter (w), which reflects relative reliance on reward ($w = 0$) versus generosity ($w = 1$) feedback. This parameter was higher—indicating greater reliance on generosity feedback—when participants were told they were interacting with other humans as opposed to slot machines. Boxes indicate interquartile range; whiskers indicate minimum and maximum points; central lines indicate the median; dots show individual data points, jittered for visualization. (b) Choice of human partners or slot machines in test phase. The plot shows the proportion of choices for which participants selected the target onscreen that was higher in generosity, and, independently, the proportion of choices for which participants selected the target that was higher in reward value. Participants relied more on generosity knowledge when told they were interacting with human partners as opposed to slot machines. The dotted line indicates chance. Error bars show standard error of the mean, with within-participants adjustment (Morey, 2008).

subsequent choices, reward and generosity information is combined using a weighting parameter w. This parameter indicates the relative influence of reward feedback and generosity feedback on choice, ranging between 0 (reliance only on reward value) and 1 (reliance only on generosity). (For full model specification and parameter fits, see Supplemental Methods and Table S1.)

We hypothesized that the balance of trait- and reward-based learning would shift toward traits when the task was framed as social as compared with nonsocial. We therefore compared the weighting parameter (w) across the two conditions using rank sum tests due to non-normality of the distribution. As hypothesized, the w parameter was higher for human targets (median = 0.57), as compared to slot machines (median = 0.26), $z = -3.78$, $p < .001$ (Fig. 2): Participants relied more on generosity information when interacting with humans than with slots, but more on reward information when interacting with slot machines than with humans. Analogous regression analyses, which fit data separately for learning and test phases, produced similar results (see Supplemental Results and Tables S2–S3).

2.2.2. Explicit evaluations

We next examined participants' reported liking of each target to test whether differences in learning led people to prefer generous humans and rewarding slot machines, relative to their respective counterparts. Ratings were submitted to a 2 (Generosity: High, Low) × 2 (Reward Value: High, Low) × 2 (Target Type: Human, Slot) repeated measures ANOVA, with Target Type as a between-participants factor (Table S4). We report partial eta-squared effect sizes with 90% confidence intervals (CIs), following Lakens (2013).

Critically, the main effects for reward and generosity were each moderated by target type (Fig. 3). First, a Reward Value × Target Type interaction, $F(1,136) = 16.76$, $p < .001$, $\eta_P^2 = 0.11$, 90% CI = [0.04, 0.20], indicated that the influence of reward value depended on target type: simple effects analysis revealed that reward value strongly influenced liking in the slot condition, $F(1,69) = 75.81$, $p < .001$, $\eta_P^2 = 0.52$, 90% CI = [0.38, 0.62], but had a relatively weaker effect on liking in the human condition, $F(1,67) = 20.55$, $p < .001$, $\eta_P^2 = 0.23$, 90% CI = [0.10, 0.36]. In other words, participants strongly preferred slot machines that provided high as opposed to low monetary rewards, but more weakly differentiated humans who provided high or low monetary rewards.

Second, and in contrast, a Generosity × Target Type interaction, $F(1,136) = 9.58$, $p = .002$, $\eta_P^2 = 0.07$, 90% CI = [0.01, 0.14], indicated a strong effect of generosity on liking in the human condition, $F(1,67) = 79.21$, $p < .001$, $\eta_P^2 = 0.54$, 90% CI = [0.40, 0.64], but a weaker effect of generosity on liking in the slot condition, $F(1,69) = 17.45$, $p < .001$, $\eta_P^2 = 0.20$, 90% CI = [0.07, 0.33]. That is, participants strongly preferred generous humans over non-generous humans but more weakly differentiated among slots machines based on generosity.

These results suggest that differences in the way people learn about human and nonhuman targets through reinforcement shape their attitudes toward each. Participants who focus more on generosity feedback during social interactions might more strongly prefer generous humans, and those who focus more on reward feedback during nonsocial interactions might more strongly prefer rewarding slots. To test this idea, we examined the correlation between participants' w parameters during the learning phase and their explicit preferences for targets varying in generosity and reward. We computed a difference score indicating the degree to which each participant's evaluations reflected the generosity versus reward value associated with targets (see Supplemental Methods). This measure was analogous to the w parameter, which indicates relatively greater reliance on generosity versus reward in learning. Across human and slot conditions, higher w scores strongly predicted preferences for targets associated with greater generosity relative to reward, $r(136) = 0.72$, $p < .001$ (Fig. 4). Thus, target-specific differences during learning—an emphasis on generosity for humans but reward for slots—were associated with different criteria for evaluating humans versus slots.

2.3. Discussion

In Study 1, participants learned more from traits than rewards when interacting with humans as opposed to slots, demonstrating that social framing influences what people learn through reinforcement. Moreover, this finding rules out the possibility that participants chose partners who shared a large proportion due solely to nonsocial learning from relative (versus absolute) rewards—for instance, adapting reward responses to contextual baselines. Instead, participants demonstrated enhanced learning from proportion feedback specifically in social interactions, consistent with the view that participants inferred trait

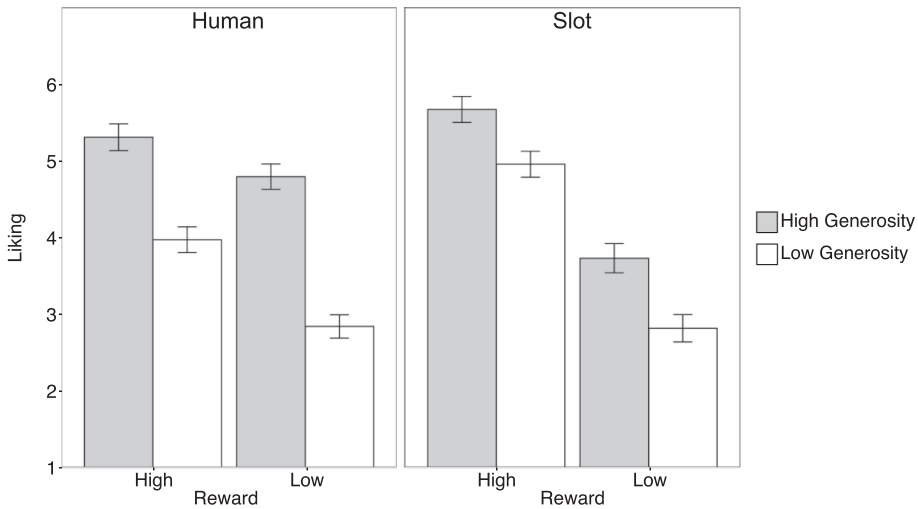

Fig. 3. The degree to which participants liked humans and slot machines in Study 1, based on the reward and generosity associated with each target during learning. Target generosity had a stronger influence on liking of humans than slots, whereas target reward value had a stronger influence on ratings of slots than humans. Error bars show standard error of the mean, adjusted for within-participant comparisons (Morey, 2008).

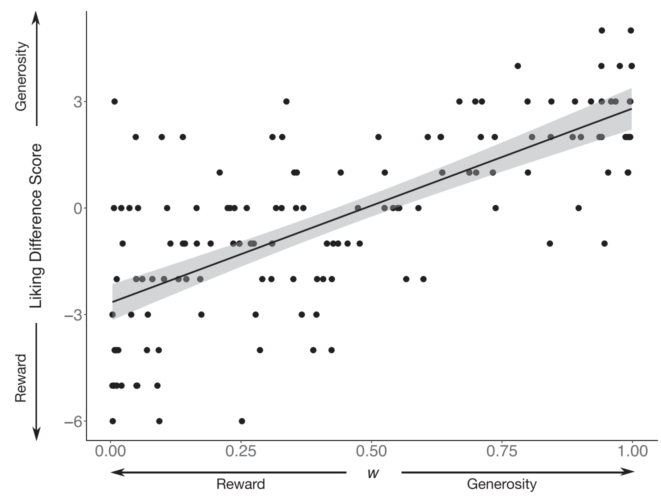

Fig. 4. Relationship between weighting parameter (*w*) during learning and reliance on generosity versus rewards in post-task evaluations in Study 1. The x-axis shows a participant's *w* parameter, indicating whether they relied more on generosity (higher values) or reward (lower values) in learning. The y-axis shows an analogous difference score indicating the extent to which a participant liked targets based on how generous they were (higher values) versus how rewarding they were (lower values). Shaded region indicates 95% confidence interval.

characteristics and valued partners with generous character. Finally, differences in learning guided explicit evaluations: participants preferred generous humans and rewarding slot machines, relative to their counterparts. These findings replicated a pilot study reported in Supplemental materials (see Figs. S1–S3, Tables S2–S3 and S5–S6).

3. Study 2

In Study 2, we aimed to replicate the findings of Study 1 while employing a fully within-participants design, allowing us to test whether differences in learning persisted even when participants encountered both humans and nonhumans. This change yielded a more stringent test of our hypothesis that learning style varies according to the social or nonsocial context, such that it may even change within participants as they shift between contexts.

3.1. Method

3.1.1. Participants

We recruited 122 individuals through Amazon's Mechanical Turk (54 female, 68 male, mean age = 33.98, SD = 9.72) who participated in exchange for payment. Sample size was determined by aiming for 100 participants, heuristically chosen to provide ample power for the within-participants design. Twenty additional participants were recruited to compensate for any necessary subject exclusions, and two additional MTurk workers completed the task without submitting their work for payment. Using the exclusion rule described in Study 1, data from 19 participants were excluded, leaving 103 participants for analysis (47 who viewed humans first, 56 who viewed slots first). We also explored robustness using the more lenient rule noted in Study 1, which excluded data from five participants. All results remained the same except for one that became stronger (see Supplemental materials). Participants were randomly assigned to complete either the human or slot block first.

3.1.2. Procedure

The procedure was similar to that of Study 1, with one exception. To allow time for two learning blocks, we omitted the test phase; instead, participants completed two consecutive blocks of learning trials (60 trials each). Although this change permitted a within-subjects test without excessive fatigue, it provided fewer responses overall and included only trials in which participants were still learning—a tradeoff yielding a more conservative test of our hypothesis. For half the participants, Block 1 involved learning from humans depicted by four fractal images and Block 2 involved learning from computerized slot machines depicted by four new fractal images. For the other half of participants, this order was reversed. Again, the role assigned to each fractal image was randomized across participants, as was the description of each image as human or slot. Participants completed evaluation ratings after each learning block and, at the end of the session, questionnaires assessing strategy and anthropomorphism. Due to an error, 11 participants saw the wrong images during some evaluation ratings and were therefore excluded from analysis of these ratings (but included in other analyses).

3.2. Results

3.2.1. Computational model

We fit data from each learning block to the computational model described in Study 1, allowing separate *w* parameters for humans and slots (see Table S7 for all parameter fits). This procedure permitted us to test for within-participant differences in *w* across target types. As hypothesized, and replicating Study 1, the w parameter was higher when learning about humans (median = 0.55) compared to slot machines (median = 0.39), $z = -2.29$, $p = .02$ (sign rank test). Again, participants learned relatively more from generosity when interacting with humans and relatively more from reward when interacting with slots machines, even though targets behaved identically (Fig. 5; see Supplemental material for regression analyses yielding similar results).

Notably, while this effect replicated the pattern observed in Study 1, it was attenuated somewhat relative to Study 1 and the pilot study (see Supplemental Results). This attenuated effect may have been due to the use of a within-participants design. Indeed, an effect of order emerged, suggesting that the target type encountered in the first block may have influenced responses to the second target (Supplemental material, Tables S8–S9). When considering only the first block viewed by each participant, Study 2 replicated the between-subjects pattern of Study 1: the w parameter was higher among participants who interacted with humans (median = 0.59) relative to those who interacted with slots (median = 0.34), $z = -3.09$, $p = .002$ (rank sum test).

3.2.2. Explicit evaluations

We again examined participants' reported liking of each target to determine whether people preferred generous humans and rewarding slot machines, relative to their respective counterparts. Ratings of liking for each target were examined in a 2 (Generosity: High, Low) × 2 (Reward Value: High, Low) × 2 (Target Type: Human, Slot) repeated measures ANOVA (Fig. 6, Table S10). As in Study 1, a Reward × Target Type interaction, $F(1, 94) = 3.94$, $p = .05$, $\eta_P^2 = 0.04$, 90% CI = [0.000004, 0.12], indicated that the effect of reward value depended on target type. Reward value strongly influenced evaluations of slot machines, $F(1,94) = 77.57$, $p < .001$, $\eta_P^2 = 0.45$, 90% CI = [0.33, 0.55], but had a relatively weaker effect on evaluations of humans, $F(1,94) = 46.89$, $p < .001$, $\eta_P^2 = 0.33$, 90% CI = [0.21, 0.44]. Although a Target Type × Generosity interaction was not statistically significant, $F(1,94) = 2.63$, $p = .11$, $\eta_P^2 = 0.03$,[1] it was qualified by a significant three-way interaction in a model that included Target Order $F(1,93) = 7.43$, $p = .008$, $\eta_P^2 = 0.07$, 90% CI = [0.01, 0.17] (Table S11, Fig. S4). Participants who first saw slot machines relied on generosity more when later evaluating humans, as in the Target Type × Generosity interaction of Study 1, $F(1,51) = 8.40$, $p = .006$, $\eta_P^2 = 0.14$, 90% CI = [0.03, 0.29]. Among participants who saw humans first, this interaction was not statistically significant, $F(1,42) = 0.93$, $p = .34$, $\eta_P^2 = 0.02$, suggesting that initial exposure to humans may lead to greater reliance on generosity feedback when subsequently interacting with slots (see Supplemental materials for more details).

Finally, as in Study 1, we tested whether individual differences in learning related to patterns of explicit evaluation. We examined the relationship between the w parameter for each target type and the difference score indicating the degree to which participants' evaluations of each target type reflected generosity versus reward (Supplemental Methods). Replicating Study 1, the w parameter during learning predicted subsequent reliance on generosity, relative to reward value, in evaluations of humans, $r(93) = 0.78$, $p < .001$, and slots, $r(93) = 0.73$, $p < .001$. Participants who learned more from generosity feedback liked generous humans more, once again linking differences in learning to differences in evaluations (Fig. S5; see also Supplemental Results for mediation analysis).

3.3. Discussion

Study 2 results replicated those of Study 1 using a within-subjects

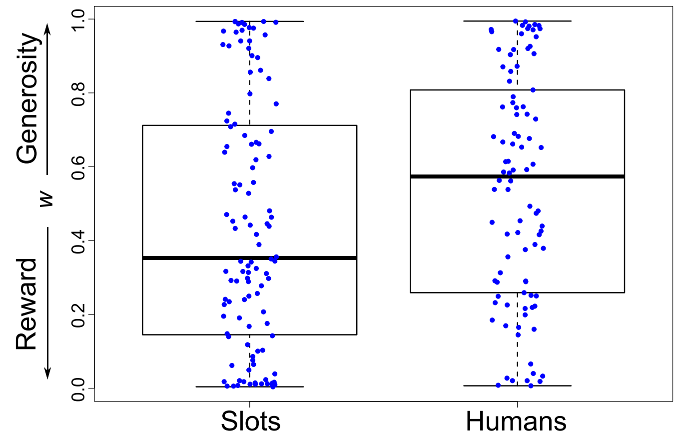

Fig. 5. Reinforcement learning from rewards as opposed to generosity in Study 2, in which participants interacted with both humans and slot machines in within-participant conditions (order counterbalanced). Learning style is indicated by a weighting parameter (w) that reflects relative reliance on reward ($w = 0$) versus generosity ($w = 1$) feedback. This parameter was higher—indicating greater reliance on generosity feedback—when participants were told they were interacting with other humans as opposed to slot machines. Boxes indicate interquartile range; whiskers indicate minimum and maximum points; central lines indicate the median; dots show individual data points, jittered for visualization.

design. This design provided a particularly stringent test of our hypothesis, given that participants completed identical tasks twice, with only one difference in the instructions, and with fewer trials. Nonetheless, a social framing still altered learning: participants relied more on generosity feedback—relative to reward feedback—when learning about humans as opposed to slot machines.

4. General discussion

Do people learn different information through reinforcement in social and nonsocial interactions? In two studies and a pilot study, we found that participants prioritized trait feedback relative to reward feedback when interacting with humans as opposed to slot machines. Specifically, they were more likely to choose interactions with generous, as opposed to rewarding, human partners. Moreover, these differences in learning predicted subsequent evaluations: participants most strongly preferred humans who were generous but slot machines that were rewarding. Thus, these findings reveal an influential role for trait inference in reinforcement learning that emerges more strongly in social contexts.

4.1. Implications for social cognition

Our findings offer theoretical implications for social cognition, learning, and attitudes. First, they demonstrate that social contexts transform the dynamics of reinforcement learning. Existing models of reinforcement learning assume that agents maximize rewards, typically indicated by their tendency to repeat choices after receiving reward feedback. Yet, we found that participants preferred interactions with generous partners, even when generosity was statistically independent of reward. Indeed, to maximize task earnings, the optimal strategy in the learning phase would have been to ignore generosity feedback and learn entirely from reward feedback (Hackel et al., 2015). Nonetheless, the social framing led people to prioritize generosity feedback over reward feedback when choosing partners. Thus, while prior work established a difference between reward learning and trait learning (Hackel et al., 2015), the present findings establish a difference between social learning and non-social learning. Crucially, our results demonstrate a unique contribution of social cognition to learning,

[1] This interaction was the only result that changed when using the more lenient exclusion rule noted in Study 1, becoming more robust, $F(1,107) = 4.90$, $p = .03$, $\eta_P^2 = 0.04$, 90% CI = [0.002, 0.12].

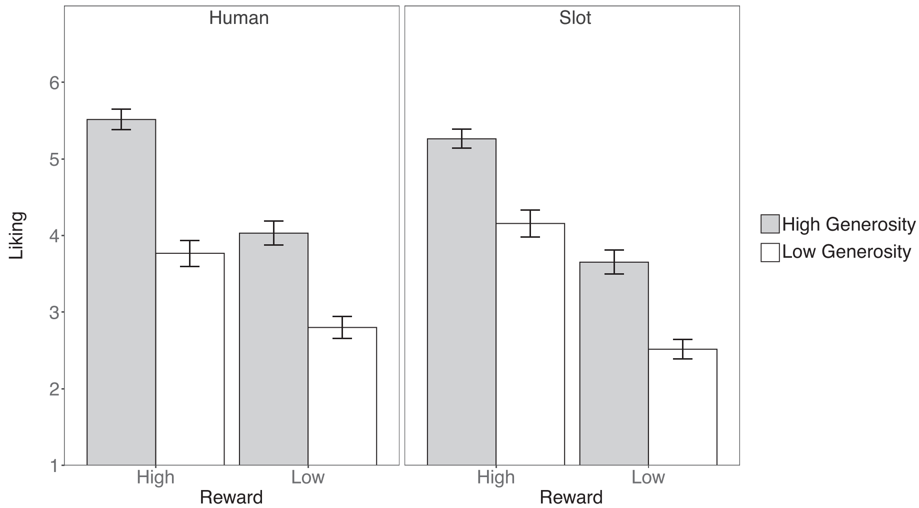

Fig. 6. The degree to which participants liked humans and slot machines in Study 2, based on the reward and generosity associated with each target during learning. Error bars show standard error of the mean, adjusted for within-participant comparisons (Morey, 2008).

above and beyond any potential nonsocial effects of relative reward.

Why did generosity feedback reinforce choice more strongly in social than non-social interactions? One possibility is that participants expected other humans—but not slots—to have stable traits across time, while expecting reward values to be unstable across time given the variation in point pools allotted to players across trials. If this were the case, the proportion of giving, rather than the absolute amounts, would provide a more stable basis for predicting future outcomes during social interactions. Although we did not cue participants to expect these task features, the assumption that trait characteristics remain stable over time is common among human perceivers (Kelley, 1973). Consistent with this idea, people especially desire fair partners rather than wealthy ones when wealth is unstable and difficult to predict across time (Raihani & Barclay, 2016), presumably because fairness represents a stable trait that can predict interaction outcomes even when wealth fluctuates. As such, other dispositional inferences—such as a person's chronic goals or values—may offer a similar basis for learning; moreover, such learning may depend on individual or cultural differences in the tendency to attribute behavior to stable internal causes (Markus & Kitayama, 1991; Morris & Peng, 1994). It is also possible that participants experienced the generosity of sharing as affectively pleasing when playing with humans, but not slot machines, consistent with the idea that social rewards can reinforce behavior (Lin et al., 2011). Each of these possibilities are consistent with our broad conclusions, and it will be interesting to explore these and other mechanisms in future research.

Trait learning was not completely unique to social interactions, however: participants encoded proportional outcomes from slots, albeit to a lesser extent. Indeed, Study 2 suggested that the learning style adopted in one context might affect learning in another: participants who first viewed humans relied more on generosity feedback when later viewing slot machines. This finding comports with past work (Hackel et al., 2015) and raises the possibility that anthropomorphism might play a role in learning (Hsu & Jenkins, 2015; see also Supplemental Analyses), as might relative reward encoding (Holroyd et al., 2004; Palminteri et al., 2015). More broadly, the observed effects of target order demonstrate an intriguing degree of malleability in learning across social and non-social contexts.

Participants also learned from rewards in social interactions—a form of learning that is well-characterized in nonsocial contexts, but typically is not featured in models of social cognition (Amodio, 2019; Amodio & Ratner, 2011; Jones, 1985). Specifically, participants chose and liked partners who previously offered rewarding outcomes, regardless of their generosity, consistent with prior work (Hackel et al., 2015; Hackel, Berg, Lindström, & Amodio, 2019; Hackel & Zaki, 2018). By dissociating modes of reinforcement learning in social and nonsocial contexts, these findings reveal distinct contributions of reward processing and trait impressions to behavior.

These results further suggest a link between reinforcement learning and attitude formation. Participants' attitudes toward humans were strongly influenced by generosity but only weakly influenced by rewards, whereas attitudes toward slot machines were not as strongly influenced by generosity. These differences in evaluation were associated with individual patterns of learning: participants who learned more from generosity feedback also liked generous targets more. Thus, it appears that participants' evaluations reflected their style of learning in addition to the objective feedback provided.

4.2. Implications for reinforcement learning

Beyond their implications for social cognition, these findings demonstrate that task framing can alter patterns of reinforcement learning and choice in a top-down manner. Past work has shown that bottom-up features of a task—such as reward statistics—alter learning in social and nonsocial contexts (Behrens, Hunt, Woolrich, & Rushworth, 2008; Behrens, Woolrich, Walton, & Rushworth, 2007). In the present studies, participants completed identical tasks, yet behavior depended on whether they believed they were interacting with humans or slot machines. Thus, reinforcement learning reflects not only environmental statistics, but also top-down beliefs. Such top-down beliefs might inform one's internal model of the environment, consistent with model-based reinforcement learning, or might alter which features of the environment feel rewarding, consistent with model-free reinforcement learning (Daw et al., 2011). Either way, reinforcement learning may vary across situations; although reinforcement learning studies often use stimuli such as slot machines, the present work suggests that real-life learning can vary based on people's expectations in a given situation.

4.3. Conclusions

To date, models of reinforcement learning have focused on the learning of reward value. Our results show that people also learn trait attributes through reinforcement, and that they do so selectively in human compared with nonhuman interactions. Indeed, the tendency to infer traits through reinforcement distinguished social from nonsocial interactions—a pattern that begins to reveal how reinforcement learning supports human social cognition.

Open practices

Materials and data for the present experiments are available at https://osf.io/gxdeu/. A pre-registration document for Study 1 is available at https://aspredicted.org/te25g.pdf.

Acknowledgments

This work was supported by the US National Science Foundation (SPRF 1606959) and the Netherlands Organization for Scientific Research (VICI 016.185.058).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2019.103948.

References

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

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science, 20*(3), 143–148.

Baetens, K., Ma, N., Steen, J., & Van Overwalle, F. (2013). Involvement of the mentalizing network in social and non-social high construal. *Social Cognitive and Affective Neuroscience, 9*(6), 817–824.

Baetens, K., Ma, N., & Van Overwalle, F. (2017). The dorsal medial prefrontal cortex is recruited by high construal of non-social stimuli. *Frontiers in Behavioral Neuroscience, 11*, 44.

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

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature, 456*(7219), 245.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214.

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*(6), 1204–1215.

Daw, N. D., O'Doherty, J., Dayan, P., & Seymour, B. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews, 37*(7), 1297–1310.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology, 11*(2), 219–224.

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.

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*, 2592.

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), 1233–1235.

Hackel, L. M., Mende-Siedlecki, P., Loken, S., & Amodio, D. M. (2019). *Context-dependent learning in social interaction: Trait impressions support flexible social choices.* (Manuscript Submitted for Publication).

Hackel, L. M., & Zaki, J. (2018). Propagation of economic inequality through reciprocity and reputation. *Psychological Science, 29*(4), 604–613.

Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39*(6), 1050.

Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.). *Person memory: The cognitive basis of social perception* (pp. 155–177). Hillsdale, NJ: Erlbaum.

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

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology, 41*(2), 245–253.

Hsu, M., & Jenkins, A. C. (2015). Character studies. *Nature Neuroscience, 18*(9), 1198–1199.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey, & E. Aaronson (Vol. Eds.), *Handbook of social psychology*(3rd ed.). *Vol. 1. Handbook of social psychology* (pp. 47–107). New York: Random House.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–292.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*(2), 107.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*.

Lin, A., Adolphs, R., & Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience, 7*(3), 247–281.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61–64.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*(6), 949.

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications, 6*, 8096.

Raihani, N. J., & Barclay, P. (2016). Exploring the trade-off between quality and fairness in human partner choice. *Royal Society Open Science, 3*(11), 160510.

Rigoli, F., Rutledge, R. B., Dayan, P., & Dolan, R. J. (2016). The influence of contextual reward statistics on risk preference. *NeuroImage, 128*, 74–84.

Rim, S., Uleman, J. S., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology, 45*(5), 1088–1097.

Riva, P., Sacchi, S., & Brambilla, M. (2015). Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied, 21*(4), 313.

Spunt, R. P., & Adolphs, R. (2015). Folk explanations of behavior: A specialized use of a domain-general mechanism. *Psychological Science, 26*(6), 724–736.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96*(1), 58.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge: MIT Press.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human?: The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113–117.

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), 237–252.