Modeling Context-Sensitive Effects of Inequity on Reinforcement Learning

Huang Ham*[1] (hamhuang@princeton.edu), Mu-Chen Wang*[2,3] (wang253@sas.upenn.edu), Joseph W. Kable[3], Adrianna C. Jenkins[3]

[1]Department of Psychology, Princeton University
[2]National Yang Ming Chiao Tung University
[3]Department of Psychology, University of Pennsylvania

*These authors contributed equally to this work.

Abstract

People are sensitive to unequal outcomes. Recent studies show that people's ability to learn action-reward mappings depends on the percentage of the reward amount they receive (compared to another person), even when such information does not change which actions are the most rewarding for the learner. However, questions remain about the degree to which this sensitivity to inequity depends on the broader inequity context. We report results from two additional experiments in which participants had the opportunity to learn action-reward mappings, with only part of the reward going to the participant. Crucially, we manipulated whether participants encountered both advantageous and disadvantageous inequity, or only one type of inequity, within the same learning block. We found that the effect of inequity differed between these two designs, suggesting that inequity affects learning in at least partly a context-dependent way. We formalized context-dependency into a new computational model which outperformed its context-independent counterpart.

Keywords: reinforcement learning; social decision-making; inequity; reward-based learning; range adaptation

Introduction

Imagine a hypothetical scenario where there are two projects you can choose to join. One project gives you 30% of its revenue, and the other gives you 90%. The rest of the revenue is given to someone else. Which project would you choose to join? Many people would probably choose to join the second one, even when the first project may potentially be more profitable overall. This example demonstrates a simple principle that people have preferences regarding how resources should be divided between themselves and others, which has been widely investigated by research in social decision-making. Overall, people display social preferences for equity and fairness by making decisions that promote equal and fair distributions of resources (Fehr & Camerer, 2007). People generally derive greater value from equal (or fair) distributions of resources than from unequal ones, but if resources are divided unequally, people generally prefer to receive the larger share. This preference arises in a variety of decision contexts and can be moderated by factors such as social distance (Strombach et al., 2015) and social group membership (Hackel et al., 2022; Jenkins et al., 2018; Kobayashi et al., 2022)

Now, imagine a second scenario in which you do not have a choice of which project to join; rather, the amount of revenue generated by the project depends directly on the actions you take: some actions generate more revenue than others. However, you would need to learn which actions generate more revenue than others through trial and error. Would your ability to do so be affected by whether you get to keep 30% versus 90% of the revenue you generate for the project? Under this scenario, the question has transitioned from a question about decision-making to a question about reinforcement learning – the ability to learn the relationship between actions and the outcomes they produce (Sutton, Barto, et al., 1998). At first glance, it seems that inequity should have no effect because learning the most rewarding action is beneficial regardless of what percentage you get to keep. Moreover, humans and other animals can often learn from rewards implicitly (Bayley et al., 2005; Cortese et al., 2020; Pessiglione et al., 2008). Despite the difference between the two scenarios, previous work has interestingly found that social inequity not only shapes decision-making but also drives reinforcement learning (Ham & Jenkins, 2025). Specifically, people learn substantially better and faster when they get to keep more than 50% of the reward (advantageous inequity) as opposed to less than 50% (disadvantageous inequity).

However, what remains unclear is how contextually flexible the effect of reward inequity is on reinforcement learning. Suppose you currently work on a project where you get to keep 30% of the revenue. Maybe 30% feels not as bad if you have already worked on another project from which you only got to keep 10% and much worse if you have already worked on one from which you got to keep 90%. Evidence suggests that human reinforcement learning can be context-dependent (Palminteri & Lebreton, 2021; Suzuki & O'Doherty, 2020), just as economic decision-making is contextual. The effect of reward on learning depends not only on the objective magnitude of the reward but also on the reference point (whether the reward is framed as loss or gain) as well as the range of possible rewards of which the person is aware. This phenomenon is known as range adaptation (Bavard et al., 2018, 2021; Molinaro & Collins, 2023). If reinforcement learning adapts to the context of the magnitude and valence of economic reward, it may also adapt to the range of possible levels of inequity. Social evaluations follow similar principles (Chang & Sanfey, 2013). Fairness is rarely judged in isolation. Instead, internal expectations create dynamic reference points for these evaluations (Vavra et al., 2018). While such context-dependency is well-documented in social decision-making, its impact on iterative reward-based learning remains unclear.

To test whether inequity affects learning in a context de-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

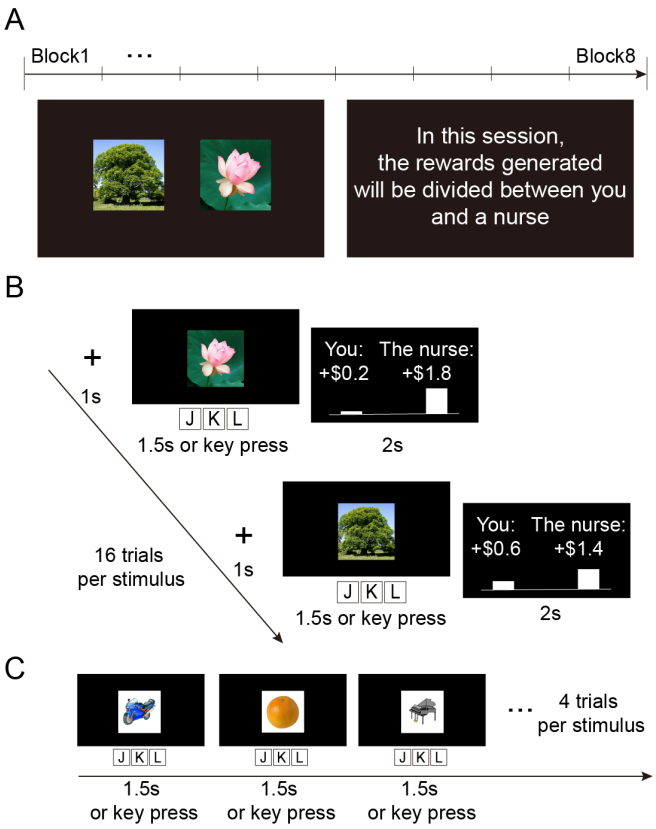

Figure 1: Experimental design: A. Block structures: The task was grouped into 8 independent blocks. Before each block, participants saw all stimuli they may encounter, and the occupation of the social target that would receive a share of the reward. B. Trial structures: Each image stimulus would appear for 16 trials in total and corresponded to one percentage of the split. The trials for all 2 stimuli were randomly shuffled. C Testing phase: All images where presented again (4 times for each image) and participants needed to recall which key generates the highest reward ($2). No feedback was given.

pendent manner, we designed two experiments in which we manipulated the inequity levels as conditions within subjects in two different ways. We found that if a participant has encountered a wider range of inequity in the recent past, the impact of inequity on learning is exaggerated. A computational model that considers context-sensitive adaptation explains the data better than one that does not. Analysis and task code for this project can be found at https://osf.io/ew5pu.

Methods

Participants

A total of 103 participants were recruited through Prolific and randomly assigned to either Experiment 1 or Experiment 2. All participants provided their informed consent prior to participation and study procedures were approved by the Institutional Review Board (IRB) of the University of Pennsylvania. Participants were prescreened on Prolific to be at least 18 years old and fluent in English. Those who failed to respond on more than 5% of trials were excluded from analysis. After applying these criteria, we retained 98 participants (Experiment 1: n=54, 29 identified as female, 25 identified as male, mean age = 34.9 years ; Experiment 2: n=44, 21 identified as female, 23 identified as male, mean age = 34.0 years).

Procedures

Participants completed a reinforcement-learning task, adapted from Ham and Jenkins (2025), which was implemented in PsychoPy (Peirce et al., 2019) and hosted online via Pavlovia. Choices led to outcomes that were split between the participant and another individual. The task consisted of eight blocks, each introducing a new set of reward contingencies. In each block, participants interacted with one stimulus set composed of two images. Across the entire experiment, there were eight distinct stimulus sets, randomly assigned to the eight blocks. Within each block, both images were presented 16 times (32 trials per block), with trial order fully randomized. Each image was associated with a fixed reward split percentage between the participant and the other person, with four possible levels: 0.1, 0.3, 0.7, and 0.9. All stimuli were images of objects adapted from multiple existing object image databases used in prior studies (Collins et al., 2017; Ham & Jenkins, 2025; Konkle & Oliva, 2012).

Experiment 1 Blocks were designated as advantageous (0.7, 0.9) or disadvantageous (0.1, 0.3), with each participant completing four blocks of each type. Block order and social target identity were randomized across participants. Social targets were indicated by an occupation label displayed at the beginning of each block to provide contextual information for the shared reward (Figure 1 A).

Every trial proceeded through the following steps (Figure 1 B). A fixation cross appeared for 1 s, after which a stimulus image was shown. The image remained on the screen until the participant pressed one of three keys (J, K, or L) or until 1.5 s had passed. For each stimulus, the three keys corresponded to three different reward amounts ($0, $1, or $2), and this key-reward mapping was randomized across stimuli. After a response, a feedback screen was presented for 2 s. The feedback displayed the amount earned by the participant and the social target, shown as bar graphs (e.g., "You: +$1.40; The nurse: +$0.60"). These values were determined by applying the reward split percentage associated with the stimulus. If participants did not respond within 1.5 s, the message "please respond faster" appeared. Responses faster than 0.15 s were not recorded, and participants were prompted to respond again.

Participants received general instructions explaining the reward structure and were informed that one randomly selected trial would determine their bonus payment. To implement the shared-reward component, one of the eight occupations was randomly selected at the end of the experiment, and the bonus from the selected trial was split equally among participants assigned to that occupation, thereby determining eligibility

for an additional bonus. The instructions did not explicitly emphasize maximizing earnings in order to avoid directing participants toward a purely self-interested strategy. Before the main task, participants completed a 16-trial practice block using a separate set of stimuli and a social target labeled "singer," which did not appear in the main experiment.

After completing all experimental blocks, participants unexpectedly completed a surprise post-test assessing their memory for the reward structure (Figure 1 C). On each trial of the post-test, a previously seen stimulus was presented for up to 1.5 s or until a response was made, after which the next trial began immediately. Unlike in the previous learning phase, participants were explicitly instructed to select the key that produced the highest reward, and no feedback was provided during the post-test. Each stimulus appeared four times in randomized order. Finally, participants completed a demographic questionnaire and rated the warmth and competence of the eight occupations that served as social targets in the experiment. Ratings were provided on separate 0–100 scales. The order of occupations and the order of the two social perception questions were fully randomized.

Experiment 2 Experiment 2 was identical to Experiment 1, except for the manner in which reward splits were configured within each block. Rather than designating entire blocks as either advantageous or disadvantageous, each block in Experiment 2 was assigned two reward split percentages, randomly drawn from 0.1, 0.3, 0.7, 0.9, allowing advantageous and disadvantageous trials to occur within the same block. At the experiment level, the sampling scheme was constrained such that each split percentage appeared an equal number of times. All other procedures and task components were consistent with those used in Experiment 1.

Model Framework

The inequity-weighted model The inequity-weighted reinforcement learning model (IRL[1]) proceeds in two stages: the value updating stage and the policy formation stage. In the value updating stage, for stimulus s and action a on trial t, the model estimates an expected value (i.e., the Q value) $Q(s_t, a_t)$ by performing an update using the delta rule (Equation 2):

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\delta_t \tag{1}$$

$$\delta_t = u_t - Q_t(s_t, a_t) \tag{2}$$

where α represents the **learning rate** and Q_t represents a $|S| \times |A|$ matrix encoding all Q values given a trial t. S represents the full stimulus/state space within a block (i.e., all the possible images that could appear). In our experiments, $|S| = 2$. A is the full action space. In our experiment, $|A| = 3$ because there were three possible buttons to press. The algorithm Q_0 is initialized as a uniform matrix of the expected values of random guessing. $\delta \in \mathbb{R}$ is the reward prediction error, and u_t is the utility received at trial t which depends on the reward $r_t \in \{0, 1, 2\}$ and the percentage of the reward given to the participant at that trial $p_t^{\text{self}} \in [0, 1]$. We formalize the effect of inequity using the Fehr-Schmidt utility function (Rohde, 2010):

$$u_t = r_t * (p_t^{\text{self}} + \gamma * (p_t^{\text{target}} - p_t^{\text{self}})) \tag{3}$$

where in our task, $p_t^{\text{target}} = 1 - p_t^{\text{self}}$ is simply the percentage of reward given to the recipient, and the **inequity weight** $\gamma \in \mathbb{R}$ controls the influence of the inequity, which we fit separately to the advantageous condition ($\gamma = \gamma_{adv}$) and the disadvantageous condition ($\gamma = \gamma_{dis}$). Q_0 is initialized to be the expected utility corresponding to the trial where a new stimulus first appears: if $\forall t' < t, s_{t'} \neq s_t$, we have $\forall a \in A$,

$$Q_0(s_t, a) = p_t^{\text{self}} + \gamma * (p_t^{\text{target}} - p_t^{\text{self}}) \tag{4}$$

In the policy formation stage, Q values are transformed by the *Softmax function* into a policy, i.e., a vector of probabilities of taking each action (represented by $\vec{\pi}_t$).

$$\vec{\pi}_t = p(\vec{A}|s_t) = \frac{e^{\beta Q_t(s_t)}}{\sum_{a \in A} e^{\beta Q_t(s_t, a)}} \tag{5}$$

where $\beta \in [0, \infty)$ represents the **inverse softmax temperature**. Finally, we allow all Q values to decay back to the initial Q values with a **decay rate** ($\phi \in [0, 1]$) to capture subjects' forgetting process.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \phi(Q_0(s_t, a_t) - Q_t(s_t, a_t)) \tag{6}$$

The context-dependent inequity-weighted model The context-dependent inequity-weighted reinforcement learning (CIRL) model extends the IRL model with three additional mechanisms: (1) a dynamic reference point that adapts to the distribution of experienced splits, (2) a compression mechanism that contracts perceived splits toward this reference point, and (3) a range adaptation mechanism that modulates sensitivity to inequity based on the range of experienced splits.

First, rather than treating splits as absolute values, the CIRL model maintains a reference point R_t that tracks the central tendency of experienced splits. The reference point updates according to:

$$R_{t+1} = R_t + \eta(p_t^{\text{self}} - R_t) \tag{7}$$

where $\eta \in [0, 1]$ is the **reference point learning rate** and $R_0 = 0.5$. Perceived splits are compressed toward the current reference point, reflecting the psychological tendency for contextual contrast effects:

$$\tilde{p}_t^{\text{self}} = p_t^{\text{self}} + c(R_t - p_t^{\text{self}}) \tag{8}$$

where $c \in [0, 1]$ is the **compression parameter**. When $c = 0$, no compression occurs; when $c = 1$, all splits are perceived as equal to the reference point.

Besides adaptation to the changing reference point, the model incorporates a range adaptation mechanism that modulates sensitivity to the inequity term based on the range of

[1]IRL does not mean inverse reinforcement learning here. We kept this acronym to be consistent with previous publications on the model.

splits encountered. We parameterize this as a range-adaptive "lift" term L_t which updates according to:

$$L_{t+1} = L_t + g(1 - L_t \cdot \Delta_t) \tag{9}$$

where $g \in [0, 1]$ is the **lift learning rate**, $\Delta_t = \max_{t' \le t}(p^{\text{self}}_{t'}) - \min_{t' \le t}(p^{\text{self}}_{t'})$ is the range of splits experienced up to trial t, and $L_0 = 1$. This mechanism causes the lift to adapt toward $1/\Delta_t$, normalizing sensitivity to the experienced range of splits. Therefore the utility function extends the Fehr-Schmidt formulation by operating on compressed splits and applying the lift selectively to disadvantageous trials:

$$u_t = r_t \cdot \begin{cases} \tilde{p}^{\text{self}}_t + \gamma_{\text{adv}}(\tilde{p}^{\text{target}}_t - \tilde{p}^{\text{self}}_t) & \text{if } p^{\text{self}}_t > 0.5 \\ \tilde{p}^{\text{self}}_t + \gamma_{\text{dis}}(\tilde{p}^{\text{target}}_t - \tilde{p}^{\text{self}}_t) \cdot L_t & \text{if } p^{\text{self}}_t \le 0.5 \end{cases} \tag{10}$$

where $\tilde{p}^{\text{target}}_t = 1 - \tilde{p}^{\text{self}}_t$. Note that the condition for selecting γ_{adv} versus γ_{dis} is based on the *original* (uncompressed) split p^{self}_t, preserving the categorical distinction between advantageous and disadvantageous conditions.

The Q-learning update rule (Equation 1), decay mechanism (Equation 6), and softmax policy (Equation 5) remain identical to the IRL model. The initial Q-values are set according to:

$$Q_0(s_t, a) = \tilde{p}^{\text{self}}_t + \gamma(\tilde{p}^{\text{target}}_t - \tilde{p}^{\text{self}}_t) \cdot \mathbf{1}[p^{\text{self}}_t \le 0.5] \cdot L_t \tag{11}$$

where $\gamma = \gamma_{\text{adv}}$ if $p^{\text{self}}_t > 0.5$ and $\gamma = \gamma_{\text{dis}}$ otherwise. $\mathbf{1}$ denotes the indicator function. In total, the CIRL model has eight free parameters: α (learning rate), β (inverse temperature), ϕ (decay rate), γ_{adv} (advantageous inequity weight), γ_{dis} (disadvantageous inequity weight), η (reference point learning rate), c (compression), and g (lift learning rate).

Modeling Procedure

We fitted all model parameters using hierarchical Bayesian methods. Compared to maximum likelihood estimation, Bayesian methods not only provide a full posterior distribution over the fitted parameters (instead of simply one point estimate), but also yield superior parameter and model recovery (Baribault & Collins, 2025; Eckstein et al., 2022). The population-level priors we used are:

$$\alpha, \eta, \lambda, \omega \sim \text{Beta}(\alpha = 1, \beta = 1)$$

$$\beta \sim \text{Gamma}(\alpha = 3, \beta = 0.5)$$

$$\phi \sim \text{Beta}(\alpha = 2, \beta = 15)$$

$$\gamma_{\text{adv}}, \gamma_{\text{dis}} \sim \text{Normal}(\mu = 0, \sigma = 1)$$

We performed fitting using the python PyMC5 package version 5.26.1 (Salvatier et al., 2016) via the no-U-Turn sampler. For each model, we ran 3 chains of 1000 tuning samples (which were discarded) and 1000 samples were kept. Therefore, 3000 samples in total were used to represent each parameter's posterior distribution, the mean of which was used as the point estimate of parameter values per participant. We simulated each model 20 times using the point estimates and compared models using the Widely Applicable Information Criterion (WAIC; Watanabe, 2013). The human data used for model fitting did not include the first iteration because the learning performance should be chance level at the first iteration and thus was not informative to model fitting.

Results

Model-free results

We first analyzed the data from the learning phase (Figure 2) using linear mixed-effects regression with the subject ID as the random intercept and the total reward amount generated per trial as the dependent variable ($0, $1, or $2). The independent variables are inequity type (advantageous or disadvantageous), iteration (1 to 16), and their interaction. We used the lmer function from the lmerTest package in R

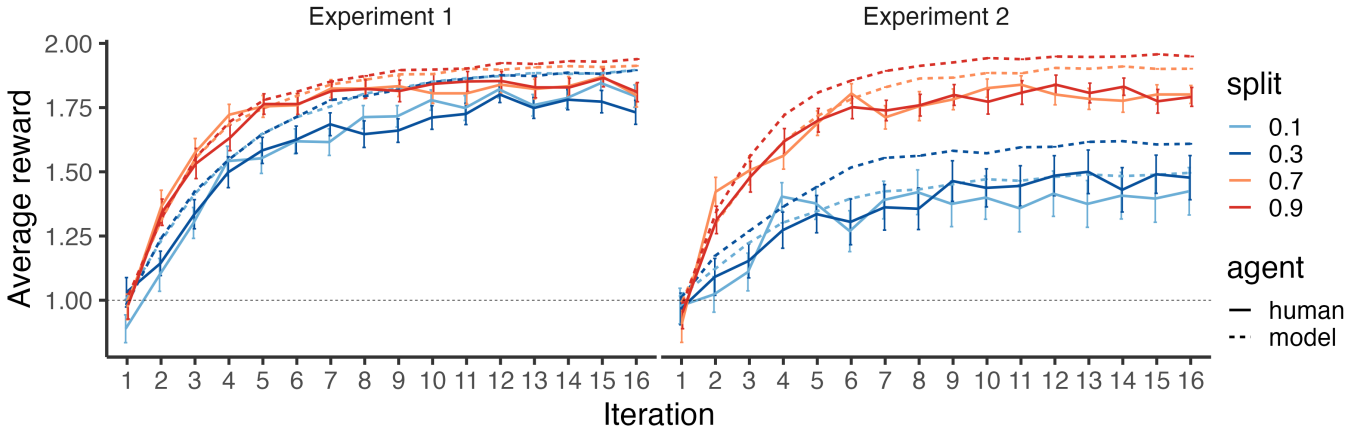

Figure 2: Learning curve: Participants overall converged on generating a higher total reward. Curves reflect the average total reward per trial as a function of the number of times that each stimulus was presented, plotted separately for each split condition (percentage of reward given to the participant). Dashed lines are the simulated learning curve by the best-fitting CIRL model. Error bars are standard error of the mean.

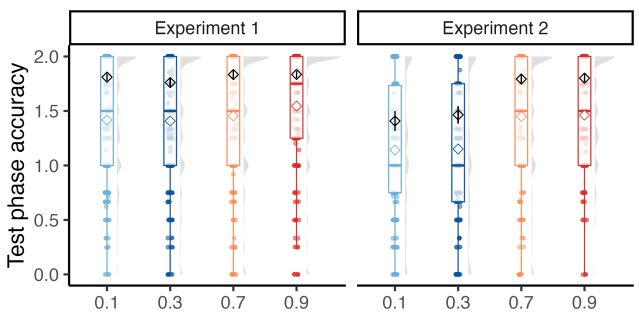

Figure 3: Testing phase: Participants were more likely to recall the most rewarding key response under advantageous inequity, especially in experiment 2. Each dot shows the mean rewardingness of the key press for each participants. Diamond reflects the population average and the middle line reflects the median. The black diamonds reflect the mean rewardingness from the last 3 iterations in the learning phase. Error bars are standard error of the mean.

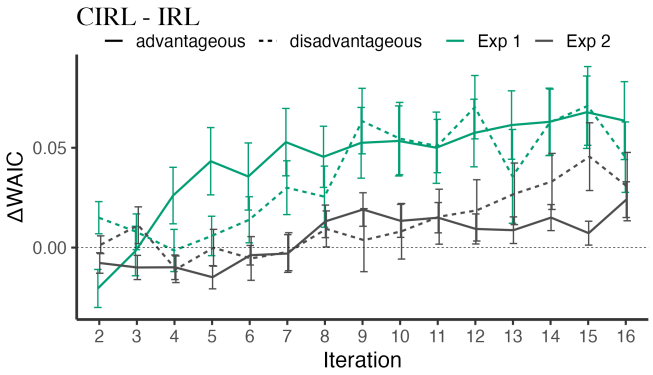

Figure 4: The trial-by-trial mean difference in WAIC between CIRL and IRL increases over the course of learning, suggesting the contextual effect enhances over learning. Error bars are standard error of the mean.

to perform regression. We found a positive effect of iteration in experiment 1 ($b = .036$, $p < 0.001$) and experiment 2 ($b = .036$, $p < 0.001$), suggesting that participants are gradually learning which key press yields a greater monetary reward over time. We also found that participants yielded a lower monetary reward overall in the disadvantageous condition in experiment 1 ($b = -.205$, $p < 0.001$) and experiment 2 ($b = -.247, p < 0.001$), replicating previous findings that participants learn better when they receive more than 50% of the reward. We further found significant interaction effects but in opposite directions in both experiments. In experiment 1, the effect of inequity decreases over iterations ($b = .010$, $p < 0.001$) but in experiment 2, the effect became stronger ($b = -.012$, $p < 0.001$). We next examined the effect of split percentage using another regression analysis with experiment ID, split percentage (0.1, 0.3, 0.7, or 0.9), iteration, and all interactions as regressors for advantageous and disadvantageous trials respectively. We found no main effect of split percentage under advantageous inequity ($p = 0.753$), but we found a positive effect under disadvantageous inequity ($b = .402$, $p = 0.017$), replicating previous findings. However, this effect seems to reduce over iterations as indicated by the negative interaction effect ($b = -.048$, $p = 0.005$).

We next analyzed whether similar effects hold in the testing phase (Figure 3) where we explicitly instructed participants to ignore the inequity information. We averaged the total reward amount corresponding to participant's responses to each stimulus in the testing phase and predicted it using inequity type and experiment ID in a linear mixed-effects model. We found that the effect of inequity remained ($b = -.093$, $p = 0.014$), suggesting that disadvantageous inequity genuinely impaired learning. Moreover, this effect is stronger in experiment 2 ($b = -.215$, $p < 0.001$). This pattern in the testing phase mirrors participant's performance at the end of the learning phase (calculated as the averaged reward amount generated by participants at the last 3 iterations of each stimulus), where we found a significant effect of inequity type, although only in experiment 2 ($b = -.312$, $p < 0.001$). The fact that these results still stand in the testing phase where we unequivocally instructed participants to ignore inequity further strengthens the claim that social inequity genuinely impairs learning.

We also found similar patterns in the reaction time data, using similar linear mixed-effects regression models with reaction time as the dependent variable. We first found a negative interaction effect in experiment 1 ($b = -.006$, $p < 0.001$) and experiment 2 ($b = -.005$, $p < 0.001$), confirming that participants responded faster in later trials because they learned the reward contingencies better. We also found that reaction times were higher overall in the disadvantageous condition in experiment 1 ($b = .018$, $p = 0.015$) and experiment 2 ($b = .024$, $p = 0.003$), suggesting worse learning in the disadvantageous condition. In the testing phase, however, we only found a significant effect on reaction time in experiment 2 ($b = .033$, $p = 0.003$), with slower reaction times in the disadvantageous condition. However, the direction of the effect is the same in experiment 1 ($b = .015$, $p = 0.103$).

Modeling results

We found that the CIRL model out-performed the IRL model in both inequity conditions and experiments ($ts > 2, ps < 0.008$). This suggests that a model that assumes context sensitivity of inequity outperforms a baseline model that assumes inequity having a fixed effect on learning. The CIRL model also outperforms the truncated versions where we fix either η, c, or g to be 0 ($ts > 3.7, ps < 0.001$), indicating that all these additional mechanisms included in the CIRL contributed to explaining the data. Moreover, CIRL model outperforms IRL more in later iterations of learning. To show this, we predicted the difference in WAIC (Figure 4) using experiment ID, inequity type, iteration, and their interactions in a linear mixed-effects regression. We found a significant positive effect of iteration ($b = .005$, $p < 0.001$) and this effect was

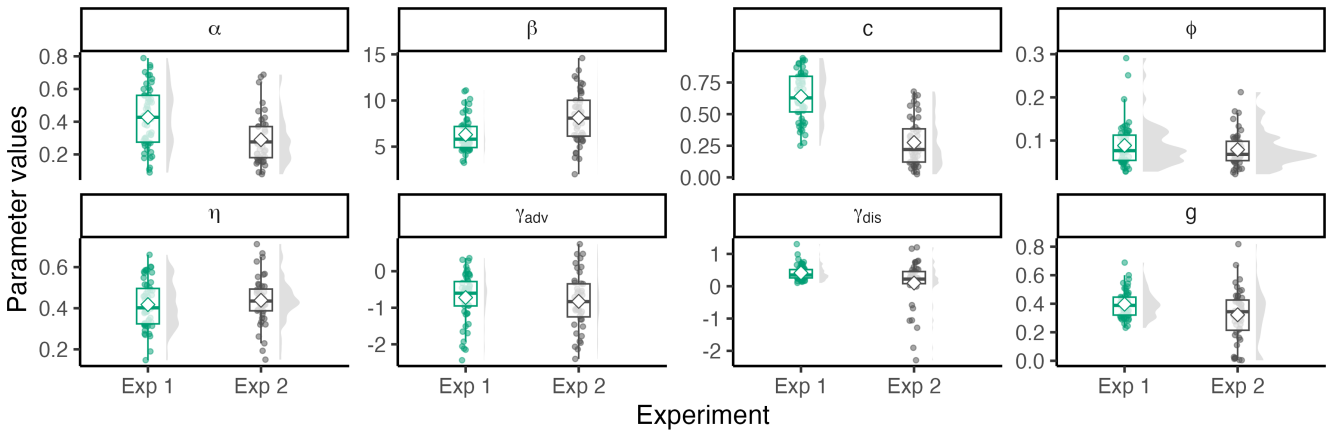

Figure 5: Parameters of CIRL: Each dot shows the fitted posterior mean for each participant. Diamonds reflect the population average and the middle line reflects the median. Error bars are standard error of the mean.

stronger in experiment 1 ($b = -.003$, $p = 0.010$). This is consistent with the finding how range adaptation exaggerates over the course of learning (Bavard et al., 2018). Simulating the CIRL model (Figure 2) shows that it captures the main behavioral patterns, especially where the inequity effect shrinks over learning in experiment 1 ($b = 0.005$, $p < 0.001$) but grows over learning in experiment 2 ($b = -0.016$, $p < 0.001$). While the IRL also produces the same pattern in experiment 2 ($b = -0.019$, $p < 0.001$), it does not show the same negative interaction effect in experiment 1 ($b = -0.001$, $p = 0.253$). Lastly, we compared the fitted model parameters of the CIRL model in both experiments using Wilcoxon test. We found that participants in experiment 1 overall had higher learning rate ($U = 1716$, $p < 0.001$), lower inverse softmax temperature ($U = 696$, $p < 0.001$), higher γ only for the disadvantageous condition ($U = 1602$, $p = 0.003$), higher compression ($U = 2144$, $p < 0.001$), and higher lift learning rate ($U = 1486$, $p = 0.034$).

Discussion

Inequity shapes how people value distributions of resources, which in turn impacts not only decision-making, but also reward-based learning. In this paper, we show that the effect of inequity on reward learning is not fixed, but dependent on the surrounding inequity context. Across two experiments, we manipulated the inequity context in different ways: in one, participants always encountered only one type of inequity within each learning block; in the other, inequity types were intermixed. We found that the effect of inequity on learning progresses differently in these two contexts. We further designed a computational model that implements context-dependent mechanisms and showed that it outperforms the original inequity-weighted reinforcement learning model.

A notable contribution of these experiments is the inclusion of a testing phase where participants were explicitly instructed to recall the most rewarding options, regardless of the inequity. This serves as a direct test of the outcome of learning, addressing one limitation of Ham and Jenkins (2025). We found the same effect of inequity on participants' performance in the testing phase, showing that social inequity fundamentally impacts one's ability to learn, instead of simply biasing the choices during the learning phase.

One limitation of the current studies is that the inequity manipulation is confounded by the reward magnitude given to the participant. While we showed in our prior studies (Ham & Jenkins, 2025) that the inequity effect held as strongly when this confound is not present, better control is needed to ascertain whether the contextual effects hold specifically for the variations in social inequity. Another unresolved question is why social inequity impacts reinforcement learning in this way. One possibility is that disadvantageous inequity demotivates participants from exerting cognitive effort to learn. Another possibility is that participants implicitly represent disadvantageously unequal options as less rewarding, despite trying to learn from them with similar effort. Future studies can focus on teasing apart these two mechanisms, either through behavioral studies or neuroimaging.

While our behavioral and computational results point toward a context-sensitive effect of inequity on learning, the neural implementation of such sensitivity remains an open question. Future fMRI studies could test predictions of our CIRL model on brain activity in the reward system. Specifically, given its role in representing reward prediction errors, the ventral striatum may show attenuated RPE signaling in response to disadvantageous inequity. Furthermore, drawing upon the framework (Chang & Sanfey, 2013), the anterior insula and anterior cingulate cortex are likely candidates for tracking dynamic reference points (R_t) and social expectations. That the range adaptive lift (L_t) is implemented through a range-scaling of neural responses in the ventromedial prefrontal cortex is another compelling possibility. Characterizing the degree to which activity in these brain regions is predicted by the model would provide significant insight into the biological plausibility of the observed context-dependency.

Acknowledgments

This research was supported by NSF CAREER award 2339853 to A.C.J. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Baribault, B., & Collins, A. G. E. (2025). Troubleshooting Bayesian cognitive models [Epub 2023 Mar 27]. *Psychological Methods*, *30*(1), 128–154. https://doi.org/10.1037/met0000554

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature communications*, *9*(1), 4503.

Bavard, S., Rustichini, A., & Palminteri, S. (2021). Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances*, *7*(14), eabe0340.

Bayley, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature*, *436*(7050), 550–553.

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: Neural computations underlying the use of social norms in decision-making. *Social cognitive and affective neuroscience*, *8*(3), 277–284.

Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *Journal of Neuroscience*, *37*(17), 4332–4342.

Cortese, A., Lau, H., & Kawato, M. (2020). Unconscious reinforcement learning of hidden brain states supported by confidence. *Nature communications*, *11*(1), 4429.

Eckstein, M. K., Master, S. L., Dahl, R. E., Wilbrecht, L., & Collins, A. G. (2022). Reinforcement learning and bayesian inference provide complementary models for the unique advantage of adolescents in stochastic reversal. *Developmental Cognitive Neuroscience*, *55*, 101106.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in cognitive sciences*, *11*(10), 419–427.

Hackel, L. M., Mende-Siedlecki, P., Loken, S., & Amodio, D. M. (2022). Context-dependent learning in social interaction: Trait impressions support flexible social choices. *Journal of personality and social psychology*, *123*(4), 655.

Ham, H., & Jenkins, A. C. (2025). Social inequity disrupts reward-based learning. *Communications Psychology*, *3*(1), 125.

Jenkins, A. C., Karashchuk, P., Zhu, L., & Hsu, M. (2018). Predicting human behavior toward members of different social groups. *Proceedings of the National Academy of Sciences*, *115*(39), 9696–9701.

Kobayashi, K., Kable, J. W., Hsu, M., & Jenkins, A. C. (2022). Neural representations of others' traits predict social decisions. *Proceedings of the National Academy of Sciences*, *119*(22), e2116944119.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124.

Molinaro, G., & Collins, A. G. (2023). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLoS Biology*, *21*(7), e3002201.

Palminteri, S., & Lebreton, M. (2021). Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences*, *41*, 144–151.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, *51*(1), 195–203.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, *59*(4), 561–567.

Rohde, K. I. (2010). A preference foundation for fehr and schmidt's model of inequity aversion. *Social Choice and Welfare*, *34*, 537–547.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3 [Publisher: PeerJ Inc.]. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, *112*(5), 1619–1624.

Sutton, R. S., Barto, A. G., et al. (1998). Introduction to reinforcement learning, vol. 135.

Suzuki, S., & O'Doherty, J. P. (2020). Breaking human social decision making into multiple components and then putting them together again. *Cortex*, *127*, 221–230.

Vavra, P., Chang, L. J., & Sanfey, A. G. (2018). Expectations in the ultimatum game: Distinct effects of mean and variance of expected offers. *Frontiers in psychology*, *9*, 992.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, *14*(27), 867–897.