Flexibly biased learning rates in social learning

Alexandra Witt[1] (alexandra.witt@gmx.net), Stefano Palminteri[2] & Charley M. Wu[1]
[1] Human and Machine Cognition Lab, University of Tübingen, Tübingen, Germany
[2] Departement d'Études Cognitives, École Normale Supérieure, Paris, France

Abstract

Research on individual decision-making often finds a positivity bias, where people weight positive outcomes more strongly than negative ones during learning. This can be beneficial when rewards are rare, by amplifying relative value differences. Yet, we know very little about learning rate biases in social settings, where a key advantage is being able to vicariously learn from the negative experiences of others. This would imply a benefit for focusing on negative outcomes when learning socially, but is at odds with the seemingly inflexible positivity bias found in individual learning. Here, we examine learning rate biases across both individual and social settings, testing for adaptivity versus generally stable biases. Overall, participants appear more flexible in their learning rate biases when learning socially than when learning individually. This implies that human social learning may be more flexible and closer to normatively optimal behavior than individual learning.

Keywords: social learning; computational modelling; positivity bias;

Introduction

In research on individual decision-making, we often find a *positivity bias*, defined by a tendency to preferentially learn from positive outcomes of our actions (Palminteri & Lebreton, 2022). In a reinforcement learning framework, this can be formalized as having different learning rates α for positive and negative outcomes, with $\alpha^+ > \alpha^-$ defining a positivity biased agent (Fig. 1a). Such a bias can be beneficial, particularly in environments with low reward probabilities. In these settings, it may be helpful to overestimate the value of rare, positive outcomes to allow for more stable learning (Fig. 1b; Lefebvre, Summerfield, & Bogacz, 2022; Cazé & van der Meer, 2013; Hoxha, Sperber, & Palminteri, 2024). The opposite is true for rich environments, where it can be helpful to overweight negative events, since they are rarer and thus more indicative of the worse option. A great deal of previous research not only commonly finds a positivity bias in participants (Eil & Rao, 2011; Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde, & Palminteri, 2017; Sharot, Korn, & Dolan, 2011; Kahnt et al., 2009; Aberg, Doell, & Schwartz, 2016; Chase et al., 2010), but also shows that this positivity bias appears to be stable, regardless of whether it is beneficial in a given condition (Palminteri, Lefebvre, Kilford, & Blakemore, 2017; Lefebvre et al., 2017; Garrett & Daw, 2020).

While learning rate biases have been extensively investigated in individual decision-making, we know little about learning rate biases when learning from others (hereafter *social learning*). This raises the question: does the positivity bias observed in individual learning extend to social contexts, or do social settings elicit a different bias—or perhaps no bias at all? Given how reliably a positivity bias is found in individual learning, it would be natural to assume that it should also hold in social settings. Additionally, simulation studies have shown that a positivity or confirmation bias may be beneficial in group settings where every member shares the same goal (Bergerot, Barfuss, & Romanczuk, 2024; Gabriel & O'Connor, 2024).

However, one of the most salient advantages of social learning is that it lets us avoid taking personal risks during learning (Hoppitt & Laland, 2013): if we observe someone eating a poisonous mushroom and subsequently dying, we do not have to try the mushroom ourselves and suffer the consequences. Within this framing, one might expect a negativity bias in social settings in order to better learn from others what to avoid. This is at odds with the seemingly stable positivity bias found in individual learning: we cannot simultaneously be both positivity- and negativity-biased. A study using ecological data from heart surgeries supports the idea of a potential negativity bias in social learning: surgeons learned more from their own successes than failures (individual positivity bias), but more from the failures than successes of others (social negativity bias; KC, Staats, & Gino, 2013).

Finally, while no empirical studies on learning rate biases in social settings have been conducted yet, some findings from the individual learning literature may also offer insight. A previous study found that there was no bias associated with learning from the results of forced choices as opposed to freely made choices (Chambon et al., 2020). Given that observational learning also involves learning about actions and their outcomes without individual control, there might similarly be no bias in social learning.

Taken together, the literature supports three competing hypotheses regarding learning rate biases in social learning. There is reason to assume that humans could exhibit a positivity bias, a negativity bias, or no bias at all when learning socially. However, the direct empirical evidence needed to determine which of these hypotheses applies—or whether the bias varies across contexts—is currently lacking. A better understanding of whether a positivity bias also extends to social settings will be informative in a multitude of ways: be-

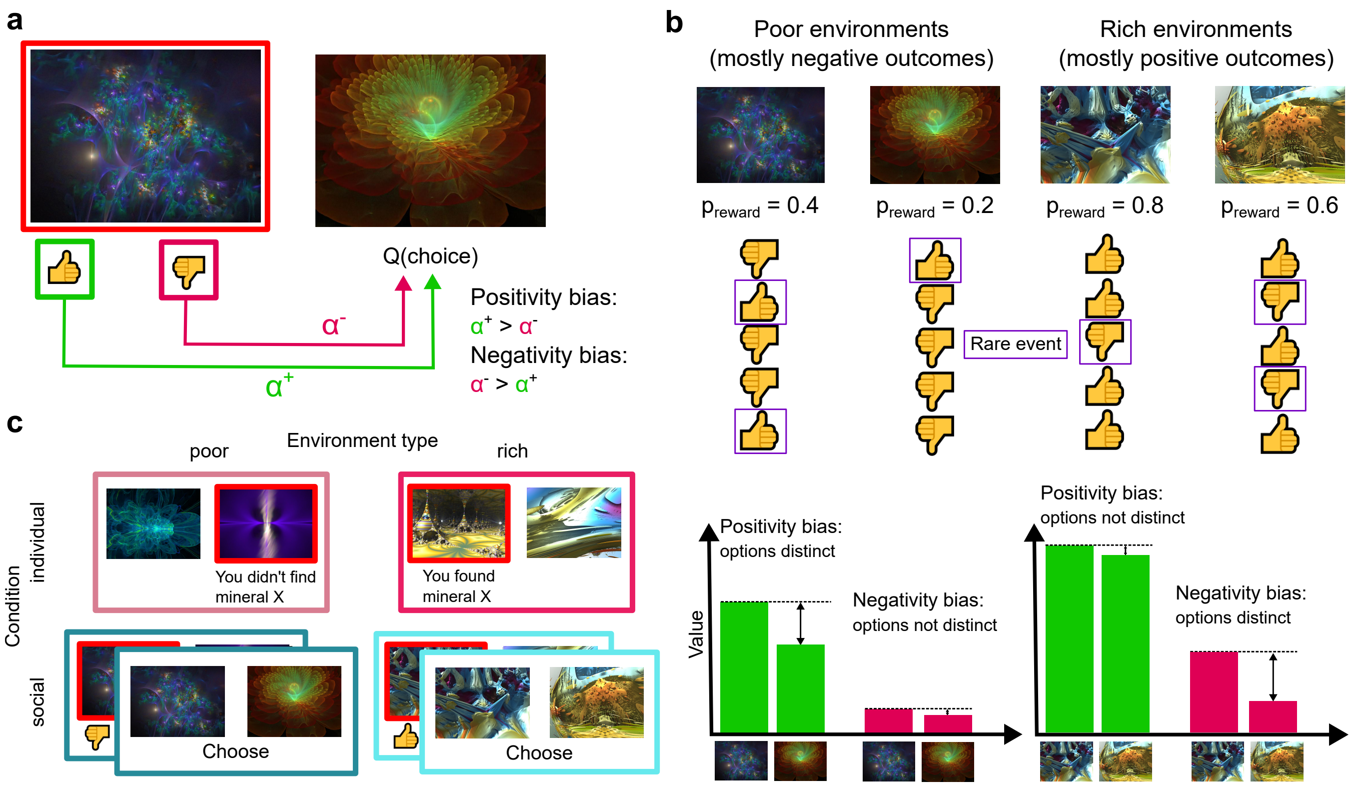

Figure 1: **Theoretical background and experiment. a)** Learning rate biases are defined as different learning rates for positive α^+ and negative α^- outcomes. **b)** Adaptivity of biases depending on environmental richness. A bias to overweighting rare events leads more distinct value estimates when comparing two options. **c)** Experiment design. We used a 2x2 within-subjects design, varying environmental richness and social condition.

sides providing insights into mechanisms of social learning, it will also let us know whether this bias—generally found to be inflexible— functions the same outside of individual settings.

To close this gap in the literature, we investigated whether a positivity bias is also present in social learning, and if so, how it changes across individual vs. social conditions and across rich vs. poor environments using a 2-armed bandit task. In the following, we first introduce the task and models, and then use simulations to demonstrate the best normative behavior in each condition. Finally, we analyze and fit models to participant data obtained from an online experiment. This allows us to compare and contrast descriptive and normative results to evaluate whether humans are able to flexibly adjust their learning rates in an adaptive manner.

Methods

To investigate learning rate biases in social settings, we varied the environment type and the learning condition. In each case, the task was a 2-armed bandit with binary outcomes.

Participants and design

We recruited *N=97* participants from Prolific (mean age: 37.6 ± 1.3 SEM; 47 female). The study was approved by the Ethics in Psychological Research Commission of the University of Tübingen (Wu_2021/0124/213), and participants provided informed consent prior to participation. On average, participants spent 36.1 ± 2.2 minutes on the task and earned £ 8.37 ± 0.12.

The study used a 2x2 within subjects experiment (Fig. 1c), varying environment type (poor vs. rich) and social condition (individual vs. social). The order of environment types were counter-balanced between-subjects, while the individual condition always preceded the social condition to ensure participants understood the perspective of the demonstrator.

The goal of the task was to maximize the amount of mineral X harvested by choosing the better of two mining sites. In *rich* environments, the reward probability of the good site p_{good} was sampled uniformly between 0.9 and 0.99. The reward probability of other option was then defined as $p_{bad} = p_{good} - 0.2$. In *poor* environments, p_{good} was sampled from a uniform distribution $U[0.26, 0.45]$, and the bad option was defined as $p_{bad} = p_{good} - 0.25$. These distributions were chosen to ensure participants could score significantly higher than chance level (based on simulation results), while manipulating what type of learning rate bias is beneficial in which condition. All participants explored the same bandits in randomized order.

In the *individual* learning condition, participants received direct feedback after each choice. For each environmental condition, there were 5 blocks of 16 trials each. Between blocks, stimulus images and reward probabilities changed. In the *social* condition, participants first observed a demonstrator make 8 choices and give reviews on each. After the observation phase, participants made one choice before moving on to a new block. This allows us to simulate sequential social learning updates without any influence of individual learning, while also roughly aligning with how humans learn from reviews (viewing multiple before making a choice). The demonstrators were defined as biased Q-learning agents, with uniformly sampled learning rates $\in [0,1]$, and a softmax temperature of $\tau = 0.3$ based on simulations. The demonstrator provided binary reviews based on the sign of their prediction error on a given trial: if the outcome received was better than expected, the demonstrator gave a positive review, and vice versa. It is worth noting that the reviews always matched the actual outcome of the bandit. Thus, both conditions share the same reward characteristics, with only the framing being changed.

Materials and procedure

The experiment was conducted online. After giving informed consent, participants received instructions about the bandits, and the nature of individual and social rounds. They were informed that their goal was to mine at the better site to maximize the collection of mineral X. In individual rounds, they used individual exploration to identify the better site. In social rounds, they relied on the demonstrator's reviews to make their choices. They were informed that there was no competition with the demonstrator, and options would not deplete with repeated visits. They did not receive any information on how the demonstrator would provide reviews, only that they were well-meaning.

Participants received feedback on their performance, as well as the corresponding update to their bonus payment after each block in the individual condition, and after all blocks in the social condition. After completing the experimental task, participants were asked for demographic information before completing the experiment.

Computational model

We assume a standard Q-learning model with different learning rates depending on the sign of the prediction error:

$$Q_{t+1} = Q_t + \begin{cases} \alpha^{+} PE_t \text{ if } PE_t > 0 \\ \alpha^{-} PE_t \text{ if } PE_t < 0 \end{cases} \tag{1}$$

where $PE_t = R_t - Q_t$ is the prediction error between the observed outcome R_t and the expected outcome Q_t. For an unbiased model, we assume $\alpha^{+} = \alpha^{-}$. We set $R_t = 1$ if mineral X was found, and $R_t = 0$ if it was not. In social learning trials, the review received from the demonstrator is based on their prediction error, so we treat positive reviews as $R_t = 1$ and negative reviews as $R_t = 0$. We then use a softmax policy $\pi(\mathbf{x}) \propto \exp(Q_t)/\tau)$ to transform the value function into choice probabilities.

Results

We first use simulations to describe the normative effects of learning rate biases in our task setting, before presenting our experimental and modeling results.

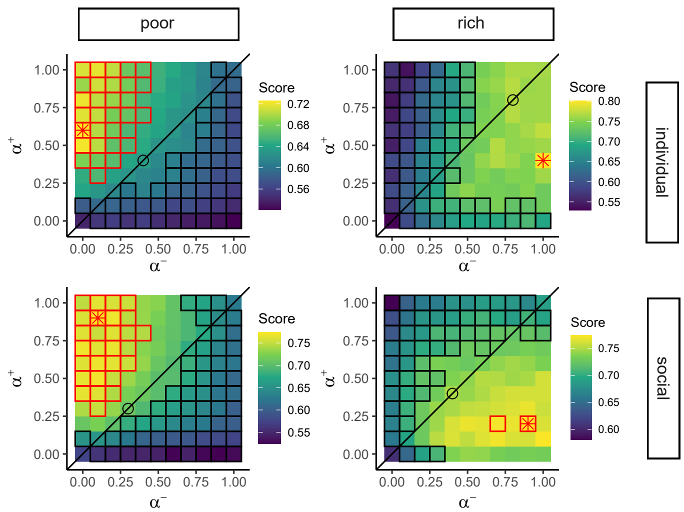

Figure 2: **Simulation results.** Performance of learning rate pairs across environments (columns) and conditions (rows). Diagonal black line shows performance of unbiased agents, with the black circles marking the best performing unbiased learning rates. Red borders indicate scores significantly higher ($p < .05$, Bonferroni corrected) than the best unbiased learner, while the black borders indicate scores significantly lower. Red asterisks indicate the highest overall performance.

Simulation results

To investigate whether a learning rate bias may be adaptive in our task conditions, we simulated 20,000 agents performing the same task as given to participants (Fig. 2). To check for benefits of biased learning rates, we first determined the best unbiased result (black circles in Fig. 2). We then tested for significant differences to the unbiased performance ($p < .05$, Bonferroni corrected). Combinations of learning rates that performed significantly better than the best unbiased agent are marked with a red border, while combinations of learning rates that performed significantly worse are marked with a black border. The best performance overall is marked with a red asterisk.

Overall, we found that in poor environments, a positivity bias is beneficial (generally higher performance above the diagonal), and negativity bias is detrimental (generally lower performance below the diagonal), in both individual and social settings. In rich environments, this effect is reversed with negativity biased agents outperforming the best unbiased agent across both social and individual conditions (albeit only significantly in the social condition).

These results are consistent with previous findings showing the normative benefits of learning rate biases depend on the richness of the environment (Lefebvre et al., 2022; Hoxha et al., 2024), and demonstrate that our experimental design allows us to test for the flexibility of human learning rate biases, especially in the heretofore untested social case.

Experimental results

First, we report the results of a two-way repeated-measures ANOVA to investigate the effect of the conditions on score (i.e., percentage of choosing the better site). Participants performed significantly better in poor than in rich environments

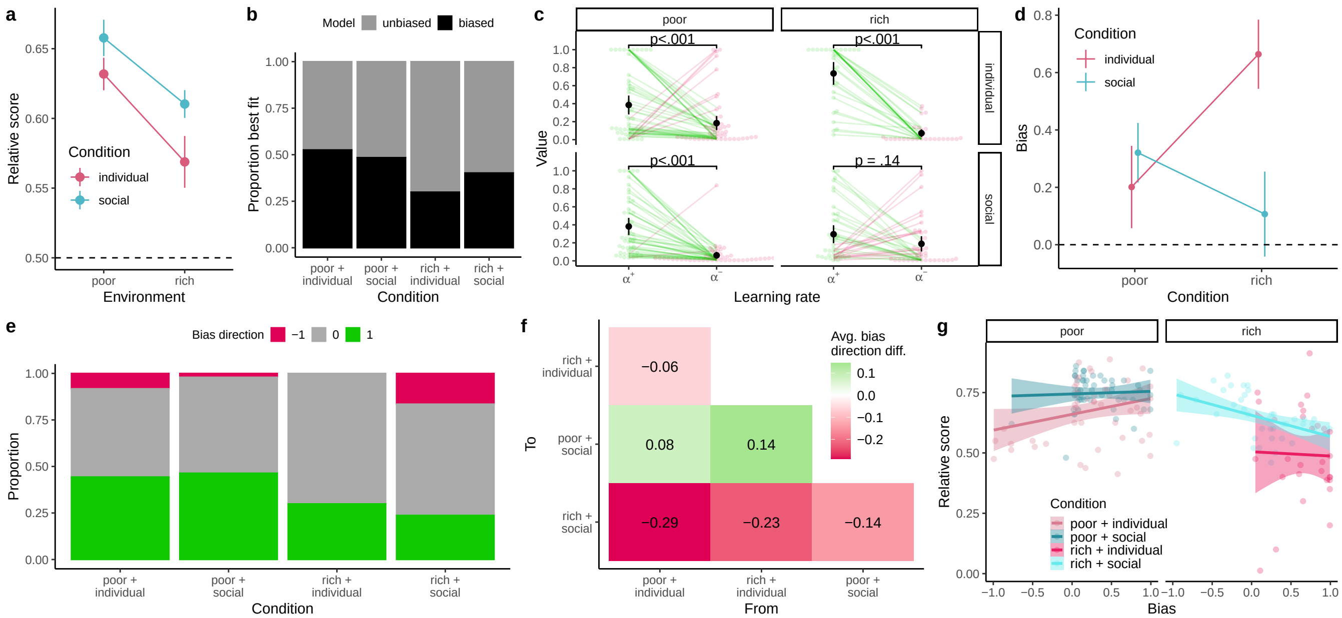

Figure 3: **Experimental results. a)** Performance across conditions. Black dashed line shows chance-level performance. **b)** Proportion of best fit model across conditions. **c)** Learning rates across conditions. Facets show conditions (environmental richness in columns, social condition in rows). Mean and standard error are shown in black. Lines connect positive and negative learning rate within participants. Green lines indicate positivity bias ($\alpha^+ > \alpha^-$), red lines indicate negativity bias. **d)** Bias across conditions. Black dashed line indicates no bias. **e)** Bias direction across conditions. Participants best fit by the unbiased Q-learning model were coded as 0. for participants best fit by the biased Q-learning model, any positivity bias was coded as 1, any negativity bias was coded as -1. **f)** Changes in bias across conditions. Coding of biases is the same way as in panel e. **g)** Participant performance as a function of individual bias.

($F_{1,96} = 16.82$, $p < .001$; Fig. 3a). This is in line with the predictions of a stable positivity bias, since a positivity bias is only adaptive in poor environments. Participants also performed slightly better in the social than in the individual condition ($F_{1,96} = 6.45$, $p = .013$; Fig. 3a) .

Model fits. We then fit Q-learning models using maximum likelihood estimation, fit separately per participant and condition, and using BIC to penalize for model complexity. The proportion of best fitting models across conditions are shown in Fig. 3b. Averaged across conditions, 42.8% of behavioral data (i.e., participant x condition) was better described by a biased than an unbiased Q-learning model, although there was some variability across conditions, with the lowest proportion of biased behavior in the rich + individual condition (29.9%), and the highest in the poor + individual condition (52.6%).

Learning rate biases. We first focus on participants best fit by the biased learning model (Fig. 3b) to analyse their learning rate biases (Fig. 3c-d only). In the biased subpopulation, we find a pervasive positivity bias in both individual conditions (Wilcoxon signed-rank test: poor + individual: $Z = -3.1$, $p < .001$; rich + individual: $Z = -4.6$, $p < .001$). However, in the social conditions, we only find a significant positivity bias in the poor environment ($Z = -5.3$, $p < .001$), in which it is adaptive, while we find no bias in the rich environment ($Z = -1.1$, $p = .14$; Fig. 3c).

The magnitude of this bias (i.e., difference between the learning rates) varied across conditions (Fig. 3d). Using a linear mixed model of `bias ~ condition*environment + (1|id)`, we find that bias was significantly stronger in rich environments ($b = 0.47 \pm 0.10$, $t_{120.65} = 4.77$, $p < .001$). However, this effect was moderated by a significant interaction with condition ($b = -0.68 \pm 0.13$, $t_{111.62} = -5.08$, $p < .001$). There was no significant main effect of condition ($b = 0.12 \pm 0.08$, $t_{97.25} = 1.36$, $p = .18$).

The exact magnitude of the bias was relatively unstable across conditions, even within participant. To relax the assumption of strict bias stability, we tried analyzing only bias direction (positive (1), negative (-1), or unbiased (0)) rather than magnitude. However, even when limiting the analysis to the direction of bias, the proportion of biases in the population varied across conditions ($\chi^2(6, N = 97) = 38.77$, $p < .001$; Fig. 3e).

Average bias direction changed most drastically between the rich + social and all other conditions, since it had the largest proportion of negativity biased individuals compared to all other conditions. We generally observe more positivity bias in the poor social condition compared to the other conditions (Fig. 3f). This shows that while the average bias based on learning rate difference was highest in the individual + rich condition (Fig. 3d), the poor + social condition actually had more individuals with a positivity bias, although the bias was weaker (Fig. 3e).

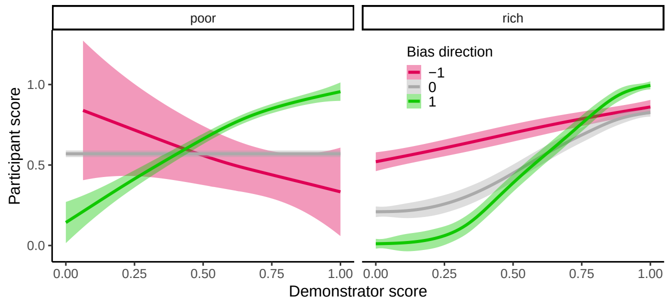

Figure 4: **Effects of demonstrator proficiency in the social condition.** Facets show environmental richness, colours indicate the direction of participant bias.

Adaptivity of biases. Next, we analyzed participant performance as a function of individual bias (Fig. 3g). In the social + rich condition, these results closely match our normative predictions: participants with a more positive bias performed significantly worse ($r_\tau = -.36$, $p = .002$; Fig. 3g, light blue line). However, we find no significant relationship between bias and score in social + poor ($r_\tau = -.02$, $p = .846$), or either individual condition (poor: $r_\tau = .12$, $p = .235$; rich: $r_\tau = -.19$, $p = .161$). This may be due to the large proportion of unbiased participants and lack of negativity biased participants in these conditions.

Table 1: Logistic regression results for score $\sim$ bias*environment + demoScore*environment + demoBias*environment + $(1|id)$

Term	OR	95% CI	p
(Intercept)	0.63	[0.49,0.82]	$< .001$
bias	2.26	[1.53,3.32]	$< .001$
rich env.	0.35	[0.27,0.46]	$< .001$
demoScore	4.34	[3.11,6.07]	$< .001$
demoBias	0.73	[0.62,0.87]	$< .001$
bias:rich env.	0.26	[0.16,0.43]	$< .001$
demoScore:rich env.	7.33	[4.95,10.85]	$< .001$
demoBias:rich env.	1.97	[1.50,2.59]	$< .001$

Social learning. Finally, we focus on the less well-understood social condition, particularly the interplay of demonstrator and participant learning. These analyses include participants best fit by the unbiased model, whose biases are set to 0. We ran a mixed effect logistical regression of `score ~ bias*environment + demoScore*environment + demoBias*environment + (1|id)`, the results of which are reported in Table 1. Overall, we find that positivity bias improves performance in poor environments, but worsens it in rich environments ($OR = 0.59$ (95 CI: $[0.42, 0.84]$), $p = .003$). This is in accordance with our normative predictions, as rich environments favor negativity biases (Fig. 1b & Fig. 2b,d).

Participants performed better with negativity-biased demonstrators in poor environments, but better with positivity-biased demonstrators in rich environments. This might appear counterintuitive since it means that a demonstrator with an adaptive bias led to worse performance. However, a demonstrator with a mismatched bias would display more switching behavior (Cazé & van der Meer, 2013), providing information about both options. This, in turn, may make it easier for participants to make better informed choices, although overall demonstrator performance would be impaired.

We also find that the demonstrator score had a significant positive effect on score, which was even stronger in rich environments. This implies that participants struggled to learn from an incompetent demonstrator's mistakes, despite the amount of information provided being identical regardless of demonstrator performance. Crucially, this does not invalidate the finding about demonstrator bias: while a demonstrator with a mismatched bias would switch more, it would also score closer to chance level. Both competent demonstrators (score close to 1) and incompetent demonstrators (score close to 0) show little switching.

To better understand this result, we investigate the effect of demonstrator score on participant score binned by participant bias (Fig. 4). While there is an overall positive trend, the relationship depends on participant bias. In poor environments, negativity-biased agents perform worse when demonstrators score higher (Fig. 4 left; red line). This is because poor environments yield few rewards, which results in largely negative reviews even when the demonstrator performs perfectly. Consequently, negativity-biased agents may falsely conclude that the chosen option must be bad, leading them to choose the inferior option. Positivity-biased agents (and unbiased agents to a lesser degree) can salvage the sparse positive reviews to make the better choice, leading to the positive relationship between demonstrator and agent score (Fig. 4 left). In contrast, rich environments yield many rewards and positive reviews, leading negativity-biased agents to successfully salvage the few negative reviews, while positivity-biased agents (and unbiased agents to a lesser degree) are misled by a large proportion of positive reviews from low-scoring demonstrators (Fig. 4, right). Thus, the strong positive effect of demonstrator score on participant performance overall is indicative of the low proportion of negativity-biased agents in our sample (Fig. 3e), rather than a general effect.

Discussion

In this study, we investigated learning rate biases in both individual and social settings. While a great deal of research has studied learning rate biases in individual learning, it is unclear whether their findings would also apply to social settings. To this end, we used a binary two-armed bandit, manipulating individual vs. social learning conditions and rich vs. poor reward environments. We first simulated behavior to ensure our manipulations cover settings in which a positivity bias is adaptive and maladaptive. Indeed, we replicate findings from prior literature showing that positivity is adaptive in poor environments, but not in rich ones. In an online experiment, we find that participants displayed a significant positivity bias in all conditions except for social + rich. This implies that social learning may be more flexible than individual learning.

Our simulations replicated the general trend of positivity biases being beneficial in poor environments and negativity biases being beneficial in rich environments (Lefebvre et al., 2022; Hoxha et al., 2024). However, there were some differences between individual and social settings. Specifically, social learning had a larger range in which a given bias was significantly better than the best unbiased estimate (Fig. 2). These differences likely arise from the difference in structure between the conditions: in the individual condition, participants repeatedly sampled options to determine which was superior, resulting in noisy behaviour. In the social condition, however, participants could integrate multiple instances of evidence from demonstrations, allowing them to make a single, well-informed decision. Despite this difference being a result of our design, we argue that it reflects real-world distinctions between individual and social learning. When learning from observation, individuals often have the opportunity to observe multiple outcomes before acting, which naturally amplifies the normative value of learning rate biases.

In the individual condition, participants deviated from these normative predictions with a higher *average* positivity bias in rich than in poor environments (Fig. 4c). However, this seems to be largely driven by a few individuals with stark differences between learning rates. The *proportion* of positivity-based individuals was actually smaller in the rich + individual condition (Fig. 4d), in line with the normative predictions. This may imply that humans are sensitive to the optimal bias to some degree, although more work on changes in learning biases in individual learning is needed to reach a robust conclusion. While prior studies often looked at learning rate biases in individual learning contexts, there are seldom comparisons between the parameters in different richness settings. Most commonly, learning rate biases are investigated in settings with equal or reciprocal reward probabilities (Palminteri et al., 2017; Lefebvre et al., 2017; Ting, Palminteri, Lebreton, & Engelmann, 2022; Kahnt et al., 2009; Aberg et al., 2016), rather than jointly high (i.e., rich environment) or low ones (i.e., poor environment). One exception is Gershman (2015), which compared low and high reward rates, and found a persistent negativity bias. This discrepancy may be caused by differences in the modelling approach or design choices (e.g., 25 trials in Gershman, 2015, vs. 80 in the current study). Thus more research is needed to better understand the sensitivity of learning biases in individual settings.

In social settings, on the other hand, participants appear closer to normative behavior. The largest proportion of participants with the normatively better bias were found in the social settings: poor + social had the most positivity-biased participants, while rich + social had the most negativity-biased. Yet, it is unclear whether this effect stems from the social framing of the task, or the requirement to observe options repeatedly before being allowed to choose. Previous literature would imply that being forced to observe choices should reduce total bias (Chambon et al., 2020), which does occur in the rich + social condition. However, this conclusion is at odds with the persistent positivity bias in the poor + social condition. This makes it more likely that the difference was caused by the social framing, although further investigation is still needed to reach a definitive conclusion.

Contrary to previous findings, we do not find biases to be stable within participant, neither in exact value, nor in direction. Prior literature has suggested that humans are fixed in their bias, be it positive or negative (Gershman, 2015; Lefebvre et al., 2017; Palminteri et al., 2017). This may be true for individual settings, in which we indeed find the greatest individual consistency out of all conditions. However, it appears that biases are much more flexible in social learning settings, where adaptation to dynamically changing social environments is a key requirement for success (Wu et al., 2025).

So far, we have framed social learning in a very positive light. However, it is worth pointing out that despite the higher flexibility of learning rate biases in the social setting, participants still performed significantly worse in rich than in poor environments. This may be due to the fact that relatively few participants showed any negativity bias, which would have been optimal in rich environments. This lack of negativity-biased participants might be caused by an inherent human tendency toward positivity bias (Palminteri & Lebreton, 2022).

While previous work has framed the positivity bias as inherently stable (e.g., for evolutionary reasons, because we have adapted to largely poor environments), our current results challenge this idea by showing biases are more adaptive, particularly in social settings. We show that participants do not persistently display a positivity bias when learning from observation, with the social setting more closely resembling the normative predictions (as also the case in Witt, Toyokawa, Lala, Gaissmaier, & Wu, 2024). Critically, the learning signal was identical across conditions due to the binary nature of the bandit, so that the only differences between conditions were the task structure (16 trials of individual trial-and-error vs. 1 choice after 8 social observations) and the framing. It seems unlikely that a change in task structure would strongly affect the direction of an otherwise stable bias, making the social framing the more likely influence on the observed biases. However, unlike previous literature, in which participants observed forced choices and displayed more unbiased learning rates overall (Chambon et al., 2020), here, we find this to be the case only in rich environments, where a positivity bias would be detrimental. In poor environments, participants remained positivity-biased, which is adaptive.

Conclusion

We replicate findings of a persistent positivity bias regardless of condition in individual learning. However, participants in the social learning condition were only positivity-biased when it was adaptive. This implies increased flexibility of learning in social compared to individual settings.

Code and data

Code and data are available at `https://github.com/AlexandraWitt/socialPositivity_CogSci`

Acknowledgments

AW and CMW are supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645, and funded by the DFG under Germany's Excellence Strategy – EXC 2117 – 422037984. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting AW.

References

Aberg, K., Doell, K., & Schwartz, S. (2016). Linking individual learning styles to approach-avoidance motivational traits and computational aspects of reinforcement learning. *PloS one*, *11*(11), e0166675.

Bergerot, C., Barfuss, W., & Romanczuk, P. (2024). Moderate confirmation bias enhances decision-making in groups of reinforcement-learning agents. *PLOS Computational Biology*, *20*(9), e1012404.

Cazé, R. D., & van der Meer, M. A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biological cybernetics*, *107*(6), 711–719.

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067–1079.

Chase, H., Frank, M., Michael, A., Bullmore, E., Sahakian, B., & Robbins, T. (2010). Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychological medicine*, *40*(3), 433–440.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–138.

Gabriel, N., & O'Connor, C. (2024). Can confirmation bias improve group learning? *Philosophy of Science*, *91*(2), 329–350.

Garrett, N., & Daw, N. D. (2020). Biased belief updating and suboptimal choice in foraging decisions. *Nature communications*, *11*(1), 3417.

Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic bulletin & review*, *22*, 1320–1327.

Hoppitt, W., & Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press.

Hoxha, I., Sperber, L., & Palminteri, S. (2024). Evolving choice hysteresis in reinforcement learning: comparing the adaptive value of positivity bias and gradual perseveration. *arXiv preprint arXiv:2410.19434*.

Kahnt, T., Park, S. Q., Cohen, M. X., Beck, A., Heinz, A., & Wrase, J. (2009). Dorsal striatal–midbrain connectivity in humans predicts how reinforcements are used to guide decisions. *Journal of cognitive neuroscience*, *21*(7), 1332–1345.

KC, D., Staats, B. R., & Gino, F. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Science*, *59*(11), 2435–2449.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 0067.

Lefebvre, G., Summerfield, C., & Bogacz, R. (2022). A normative account of confirmation bias during reinforcement learning. *Neural computation*, *34*(2), 307–337.

Palminteri, S., & Lebreton, M. (2022). The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*, *26*(7), 607–621.

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S.-J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS computational biology*, *13*(8), e1005684.

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature neuroscience*, *14*(11), 1475–1479.

Ting, C.-C., Palminteri, S., Lebreton, M., & Engelmann, J. B. (2022). The elusive effects of incidental anxiety on reinforcement-learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 619.

Witt, A., Toyokawa, W., Lala, K. N., Gaissmaier, W., & Wu, C. M. (2024). Humans flexibly integrate social information despite interindividual differences in reward. *Proceedings of the National Academy of Sciences*, *121*(39), e2404928121.

Wu, C. M., Deffner, D., Kahl, B., Meder, B., Ho, M. H., & Kurvers, R. H. (2025). Adaptive mechanisms of social and asocial learning in immersive foraging environments. *Nature Communications*, *16*, 3539.