Flexible integration of social information despite interindividual differences in reward

Alexandra Witt[1,*], Wataru Toyokawa[2,3], Kevin N. Lala[4,+], Wolfgang Gaissmaier[2], and Charley M. Wu[1]

[1]Human and Machine Cognition Lab, University of Tübingen, Tübingen, Germany
[2]Social Psychology and Decision Sciences, University of Konstanz, Konstanz, Germany
[3]RIKEN Center for Brain Science, RIKEN, Wako, Japan
[4]School of Biology, University of St Andrews, St Andrews, United Kingdom
[*]alexandra.witt@gmx.net
[+]formerly Laland

ABSTRACT

There has been much progress in understanding human social learning, including recent studies integrating social information into the reinforcement learning framework. Yet previous studies often assume identical payoffs between observer and demonstrator, overlooking the diversity of real-world interactions. We address this gap by introducing a socially correlated bandit task that accommodates payoff differences among participants, allowing for the study of social learning under more realistic conditions. Our novel Social Generalization (SG) model, tested through evolutionary simulations and two online experiments, outperforms existing models by incorporating social information into the generalization process, but treated as noisier than individual observations. Our findings suggest that human social learning is more flexible than previously believed, with the SG model indicating a potential resource-rational trade-off where social learning partially replaces individual exploration. This research highlights the flexibility of humans social learning, allowing us to integrate social information from others with different preferences, skills, or goals.

Introduction

Imagine you are in a foreign city, trying to decide on a restaurant to visit for dinner. You check reviews within a certain radius. Do you go for the best-rated restaurant no matter what, trusting the majority judgement? Or do you assume your taste may differ from everyone else's in this city, and discount ratings based on your personal preferences, integrating what is popular with what you know about your personal tastes in food? It may seem obvious that you would not generally assume that everyone you could possibly rely on for their opinion will share your exact tastes. However, much of the literature in social learning has focused on this idea of how we use information from others who are just like us[1–6].

Research on the use of social information has identified various *social learning strategies* (SLS) commonly deployed by humans[7–11]. These SLS serve to selectively limit imitation to cases in which it would be beneficial to imitate others, and can be categorized into *when*-, *what*-, and *who*-strategies. When-strategies determine when social learning should be used, e.g. when an agent is uncertain[12–14], or when individual learning is costly[14,15]. What-strategies specify what is preferentially learnt from others, e.g. emotionally evocative content[16,17], and information relevant for survival[18–20], or about social relationships[18,21]. Who-strategies determine who should be learnt from, e.g. prestigious[22,23] or successful[24–26] individuals, or the majority[1,13,14,27]. Prior research has also sought to understand *how* people imitate, that is, what mechanism underlies their use of social information, e.g. stimulus enhancement[28], decision biasing[1], or value shaping[2].

However, even when selectively limiting when and how imitation should be used for social learning, individuals may need to share the same goals or preferences as whoever they are imitating for imitation to yield favourable outcomes. In previous research, it has commonly been the case that demonstrators had the same payoff function as the participant[1–6]. Only few studies have considered social information use in matters of taste[29–31], and they have largely focused on the normative question of how best to craft social recommendations. Therefore, our understanding of the human ability to learn from others has been limited to settings in which imitation is optimal.

In real life, however, people can rarely assume that any stranger they may choose to imitate will share their exact goals. For instance, if the goal is to get home after work, following the first car in view is unlikely to lead to the desired outcome. Conversely, if an individual notices a usually bustling street deserted during rush hour, they may correctly choose to dodge the roadwork that has caused everyone's paths to change while still getting to the right house after a quick detour. This difference in exact goals (e.g. the destination of a trip) despite some shared preferences (e.g. avoiding traffic or closed roads) is commonly seen in many choice domains, like food selection, fashion, career choices, holiday planning, or scheduling, in which we can commonly learn from others. Thus, there must be more to social learning than just imitation: some consideration must be made of whether the interests of the imitating and imitated individuals are aligned.

The question of how humans learn socially from demonstrators with differing preferences is sometimes answered with Theory of Mind inference[32–34], i.e. the ability to infer others' mental states, like their goals or preferences, from their behaviour. Research in this domain has uncovered much about people's ability to infer mental-state information from others' behaviours[35–37], and specifically people's ability to infer others' preferences[38,39]. After such inference, people might indeed be able to determine whether they share another person's reward function, and thus whether exact imitation is a promising option. However, even if reward functions are not perfectly aligned, people may still be able to glean valuable insights that enhance their individual decision-making. Moreover, people in the modern world often make choices using social information that is merely an aggregate rating of others' opinions, with no way of inferring how similar each individual may be to them. Thus, there is an open question of how we can use inferred or otherwise-gained value information from others who do not share our exact preferences, which is what the current study aims to address.

Goals and scope

To this end, we introduce the *socially correlated bandit task*, which lets us investigate learning and exploration dynamics in social settings where exact imitation is not optimal. The task is based on the spatially correlated bandit[40], which uses spatially correlated rewards to allow for individual generalization, and is typically used to investigate asocial learning. We add social correlations to this setup, enabling the generalization of not only individual, but also social information (Fig. 1a-b).

In our socially correlated bandit, participants search individualized environments, which are correlated with one another. Thus, the highest rewards are generally in the same region for all participants, but directly copying another participant's best choice will not lead to the maximum payoff for oneself. This emulates the relationship of social information and diverse individual preferences and circumstances in the real world: while there are some standards that apply to everyone, not everyone would agree on the same option being optimal. While the spatially correlated multi-armed bandit has previously been used to investigate social learning, it was either in individual settings[41] or with both participants in the same environment[42], not with correlated rewards across participants.

Participants explored these socially correlated environments in groups of four. In group rounds, they had full information about other participant's choices and outcomes (Fig. 1c), thus sidestepping the actual social inference. We ran evolutionary simulations with multiple candidate models to find the normatively best strategy, which was our novel "Social Generalization" (henceforth "SG") model. We then fit these models to the behavioural data collected in two online experiments. In Exp. 1, which consisted only of group rounds, we studied whether humans would be able to utilize social information in this novel setting, and if so, how they integrated it into their decision-making. We found that participants were able to use social information to their benefit, with search behaviour being significantly influenced by other participants finding high rewards. Their behaviour was most accurately predicted by SG. We then ran Exp. 2 as a preregistered replication[43] interleaving solo rounds and group rounds. This allowed us to disentangle behavioural signatures stemming from the correlated task structure from actual social learning. It also let us delve deeper into differences between individual and social learning in the task by comparing baseline learning model's parameters between conditions. Again, we find adaptive use of social information, with SG being the best fit model. Differences in exploration behaviour indicate that social learning may function as an exploration mechanism when available[9]. Taken together, we find that humans can integrate social information with more nuance than what previous task designs implied, potentially using it to partially replace individual exploration.

Results

We use the socially correlated bandit (Fig. 1a) for this study. Each agent explores a multi-armed bandit arranged as a grid with spatial correlations[40], and can observe the other agents of their group doing the same. We generated sets of four positively correlated bandits (for details, see Methods), so that social information can be valuable, but is less so when used verbatim (Fig. 1b). In the experiments, this was framed as collecting salt samples in alien oceans as a team of scientists, with each scientist being interested in a different salt.

In the following, we first introduce four candidate models that differ in how they integrate social information into the reinforcement learning process (Fig. 2a, top panel). We then use these models in evolutionary simulations to find the best normative strategy. Finally, we report results from two online experiments. In Exp. 1, we investigated whether and how humans would be able to use social information in this new setup to enhance their decision-making. We expand on these results in Exp. 2 as a preregistered replication, where we interleave solo and group rounds to investigate how social learning influences individual exploration patterns.

Models

We first introduce an asocial baseline model (Asocial Learner; AS), followed by our candidate social models. We consider three social models, all of which build on the asocial baseline model. Each social model integrates social information into a different stage of the individual decision-making process: the policy (Decision Biasing; DB), value function (Value Shaping; VS), or reward generalization (Social Generalization; SG). All models are illustrated in Fig. 2a.

Asocial Learner (AS). We use a Gaussian Process Upper Confidence Bound (GP-UCB) model[40] as a commonly used[41,42,44] asocial baseline for the spatially correlated bandit problem. Gaussian Process regression is used to model

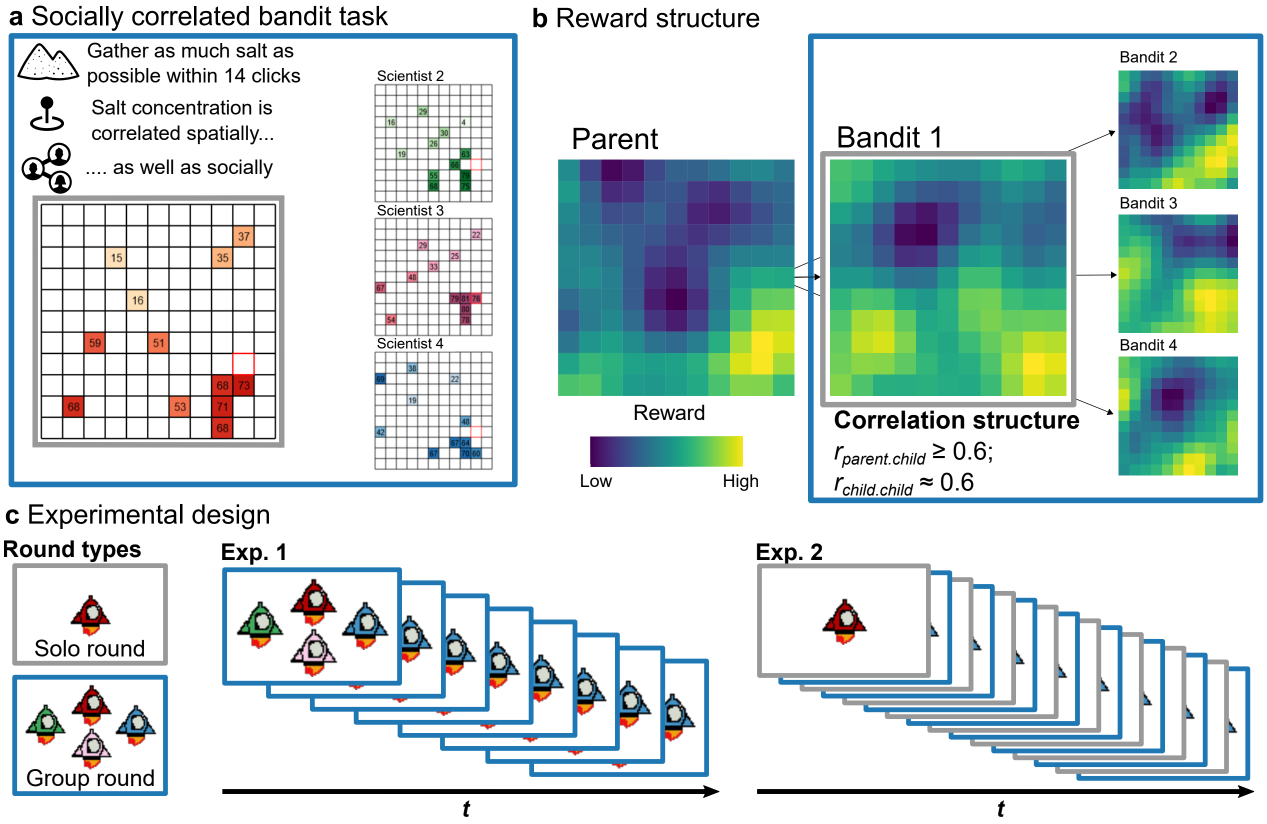

Figure 1. Experiment overview. a) Screenshot of the socially correlated bandit task. Participants completed the task either individually (solo rounds, gray border) or in groups of four (group rounds, blue border). In the group condition, they had access to choice and outcome information of other group members. Participants were instructed they would collect salt samples on alien oceans with other scientists to explain the spatial and social correlation structure. For details, see Methods. **b**) Reward structure of the socially correlated bandit. Individual payoffs are generated from a common parent grid and are positively correlated. This leads to high and low payoffs being in the same general area across participants, while global optima are still distinct, limiting the effectiveness of exact imitation. **c**) Experimental design. Exp. 1 only included group rounds, while Exp. 2 had alternating group and solo rounds, in counterbalanced order.

expectations about the reward r associated with each action by generalizing from reward observations. For some novel option $\mathbf{x}_*$ (i.e. a tile on the grid), and given past observations $\mathscr{D}_t = \{\mathbf{X}_t, \mathbf{y}_t\}$ of choices $\mathbf{x}_1, \ldots \mathbf{x}_t$ and rewards $y_1, \ldots y_t$, the posterior reward distribution is a multivariate Gaussian:

$$p(r(\mathbf{x}_*|\mathscr{D}_t) \sim \mathscr{N}\left(m(\mathbf{x}_*|\mathscr{D}_t), v(\mathbf{x}_*|\mathscr{D}_t)\right) \tag{1}$$

The posterior is thus defined by its mean $m(\mathbf{x}_*|\mathscr{D}_t)$ and variance $v(\mathbf{x}_*|\mathscr{D}_t)$:

$$\begin{aligned} m(\mathbf{x}_*|\mathscr{D}_t) &= \mathbf{K}_{*,t}^{\top}(\mathbf{K}_{t,t} + \sigma_\varepsilon^2 \mathbf{I})^{-1}\mathbf{y}_t \\ v(\mathbf{x}_*|\mathscr{D}_t) &= \mathbf{K}_{*,*} - \mathbf{K}_{*,t}^{\top}(\mathbf{K}_{t,t} + \sigma_\varepsilon^2 \mathbf{I})^{-1}\mathbf{K}_{*,t}, \end{aligned} \tag{2}$$

Here, $\mathbf{K}$ is the covariance matrix between different subsets of observations ($*$ for new inputs and t for prior observations), σ_ε^2 is the observation noise, and $\mathbf{I}$ is the identity matrix.

The assumed covariance depends on the kernel function $\mathbf{k}$, which determines how the model generalizes. We use a Radial Basis Function (RBF) kernel $k_{RBF}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\lambda^2}\right)$. The length-scale λ determines the decay rate of the covariance between two points as a function of distance, with higher values of λ assuming stronger spatial correlations. Thus, λ controls the range of generalization, with higher values leading to broader generalization, as a single data point affects more of the surrounding data. This follows the same principle as the generating function, presenting a reasonable solution to the individual generalization process. The GP also models the environment's observation noise σ_ε^2, which allows for the model to not overfit noise.

After inferring reward, upper confidence bound (UCB) sampling is used to balance exploration and exploitation tendencies. This combines posterior mean and variance resulting in a *UCB value*.

$$UCB(\mathbf{x}) = m(\mathbf{x}|\mathscr{D}_t) + \beta\sqrt{v(\mathbf{x}|\mathscr{D}_t)} \tag{3}$$

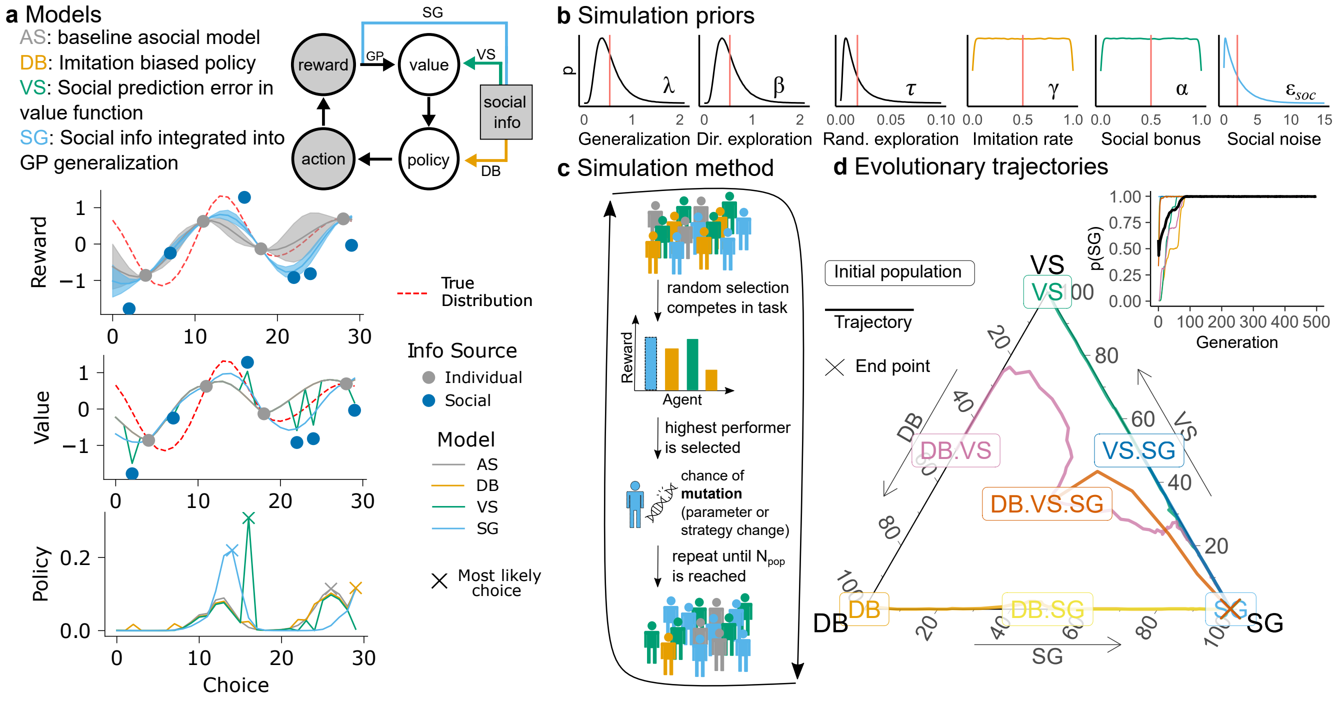

Figure 2. Models and evolutionary simulations. **a**) Model overview. Top panel: Illustration of the individual decision-making circuit and the stages at which social information is integrated. Bottom panels: An illustrative 1D example of how models incorporate social information within the steps of the reinforcement learning circuit, where the x-axis is the discrete choice space. Reward: Only SG integrates social information into the GP posterior, whereas the other models (only AS shown for ease of reading) generalize only individual information. Value: VS integrates social information into the value function proportional to its deviation from expected value. Policy: DB integrates social information into the policy based on choice frequency. Crosses mark the most likely choice for each model. **b**) Simulation priors. Prior distribution densities used for model (including evolutionary) simulations. Red line shows the mean. **c**) Evolutionary simulation method. Agents were randomly selected to compete in one round of the task. The highest scoring agent was selected for the next generation, with a chance of parameter or type mutations. **d**) Results of the evolutionary simulations. Labels show the starting point of the various initial populations. Lines show evolutionary trajectory of model proportions, with crosses at the end point after 500 generations (all at bottom right vertex). For ease of reading, the inset plot shows only the development of SG over generations.

The uncertainty-directed exploration parameter β trades off the value of an option against the uncertainty of that estimate: as it approaches 0, an agent will preferentially exploit the best known option, whereas higher β values induce more exploratory behaviour, by optimistically inflating the value of more uncertain options.

We then use a softmax to convert the value function into the *policy*:

$$\pi(\mathbf{x}) \propto exp(UCB(\mathbf{x})/\tau) \tag{4}$$

The temperature parameter τ controls how deterministically the model follows the value function: the higher it is, the more random the choices become. An agent's next action is chosen based on this policy.

Decision Biasing (DB) is the simplest social learning model, incorporating social information into the policy in a frequency-based manner[1]. This means that the choice probability for a given option is increased proportionally to how many agents have chosen that option. The policy becomes:

$$\pi = (1-\gamma)\pi_{ind} + \gamma\pi_{soc}, \tag{5}$$

with the social policy π_{soc} tracking the other agents' choices in the previous trial such that $\pi_{soc}(x) \propto n_{x_{soc},t-1}$. Here, n is the number of times an option was chosen. Individual and social policies are then combined, with the weight of social learning dependent on the mixing parameter γ.

Value Shaping (VS) incorporates social information into the value function. In previous studies, this was done by treating a social choice as a "pseudo-reward"[2]. It can be seen as an implementation of either stimulus enhancement or local enhancement[28], in this case increasing the likelihood of choosing the same option one has seen chosen by the demonstrator by increasing its value. Previous implementations of this model had no reward information, as the action outcomes were not shown in their tasks. As outcomes are shown in our task, we augment the model to be value-sensitive by using a simple prediction error approach:

$$V(\mathbf{x}) = V_{x,ind} + \alpha(V_{x,soc} - V_{x,ind}) \qquad (6)$$

with $V_{x,ind}$ being the individual UCB-value, and $V_{x,soc}$ the social value of a given option x. Thus, an observed action's value will be increased when it is better than individual expectation and decreased when it is worse. Social bonus parameter α governs the strength of this social influence. While including value information in VS improved it over a value-agnostic version (Fig. S2a), the same was not true for DB (Fig. S2b). Thus, we elected to keep using the simpler, equally good model for DB, but modified VS for better performance.

Social Generalization (SG) is a novel model that incorporates social information at the stage of the Gaussian Process regression. This means that, unlike in the other models, social information is generalized to surrounding options as well, which corresponds to a non-specific form of local enhancement[28]. However, social information is assumed to be noisier, and thus less reliable, than individual information. The formerly scalar noise term σ^2_ε (Eq. 2) becomes a vector, with its value depending on whether an observation was individual ($\delta_{\text{soc}}(\mathbf{x}) = 0$) or social ($\delta_{\text{soc}}(\mathbf{x}) = 1$).

$$\sigma^2_{\varepsilon|\mathbf{x}} = \sigma^2_{\varepsilon_{\text{ind}}} + \delta_{\text{soc}}(\mathbf{x}) \cdot \sigma^2_{\varepsilon_{\text{soc}}}, \qquad (7)$$

In addition to the term for the environment's observation noise $\sigma^2_{\varepsilon_{\text{ind}}}$, social noise $\sigma^2_{\varepsilon_{\text{soc}}}$ is added to any social observations. This social noise (henceforth referred to as ε_{soc}) determines the reliance on social information. Higher social noise causes posterior means to deviate less from the prior mean and posterior variances to remain higher in social compared to individual observations. As social noise term ε_{soc} approaches 0, social information is relied on more and more, with the extreme case of $\varepsilon_{soc} = 0$ treating social information as equally reliable as individual information.

Evolutionary simulations

We first used evolutionary simulations (see Methods) to determine which model achieves the best normative performance in this setting. Since social learning strategies have *frequency dependent fitness*[45], how well they perform depends on the frequency of other strategies in the population. Thus, evolutionary simulations (starting from different initial populations) are well-suited to evaluate frequency-based fitness without having to exhaustively evaluate every possible population composition. We considered initial populations with all possible combinations of models, with equal proportions of all included models. Simulated agents were parameterized by drawing from their model's respective prior distribution for parameters (Fig. 2b; for details see SI). Following the principle of tournament selection, we then sampled groups from the populations with replacement, selected the highest scoring agent per group for the next generation with some chance of parameter and type mutation (Fig. 2c).

We first evaluate model performance in a setup to replicate results from previous literature[2], having two agents in the same environment ($r = 1$) with one of them making optimal choices as an expert. We replicate VS being the best model compared to AS and DB in previous literature (Fig. S4a). In identical environments, VS and SG make the same predictions, and the evolutionary simulations are tied between the two models with no clear winner (average $p(SG) = .48$ and $p(VS) = 50$ in the final generation; Fig. S4b-d).

Figure 2d shows the results of evolutionary simulations of the competing social learning models in our current task environment: all possible initial populations, even ones that did not originally contain SG agents, evolve to be 100% Social Generalization agents (see Fig. S3a for starting populations including AS). This clearly suggests that SG is the normatively best model in our task.

As the parameters evolve throughout the simulations, we can also glean insight into what combinations of parameters were normatively optimal. Investigating the evolved parameters for SG (Fig. S3b), we find that λ nearly reaches the true underlying value of the environments, 2 ($\lambda = 1.96$). The random exploration parameter τ is fairly low at roughly 0.006, showing mostly deterministic choices based on the value function. The social noise parameter ε_{soc} shows considerable variation, but evolves to 3.2 on average. As this is higher than 0, we can see that indiscriminate social information use (like imitation) is not optimal in our task. The directed exploration parameter β evolves to lower values than what has previously been found in humans[40] with an average of .19. This may indicate that directed exploration can be replaced by social information use in social learning settings.

Experiment 1. People flexibly use diverse social information

Having determined SG to be the normatively best strategy for the socially correlated bandit, we now move on to online experiments using the task to see how human participants actually use social information. Exp. 1 consisted exclusively of group rounds, meaning that participants always had access to the choices and outcomes of the members of their group.

Behavioral results

Firstly, participants improved across trials (Fig. 3a), with the average performance being significantly higher than the chance level of 0.5 ($t(127) = 59.8$, $p < .001$, $d = 5.3$, $BF > 100$). There was a small but negligible learning effect over rounds (Fig. S1a). Average social search distance (the Euclidean distance between an option chosen at trial t and one chosen by another participant at t-1) decreased over time, indicating increased clustering of participant choices as the task progressed (Fig. 3b). The social search distance was also significantly lower than what would be predicted by random choice (an average of 5.75; $t(127) = -29.4$, $p < .001$, $d = 2.6$, $BF > 100$).

This social clustering may have stemmed from a tendency to approach other participants who have earned a high reward in the previous trial. A regression of search distance over previous reward by information source (individual or

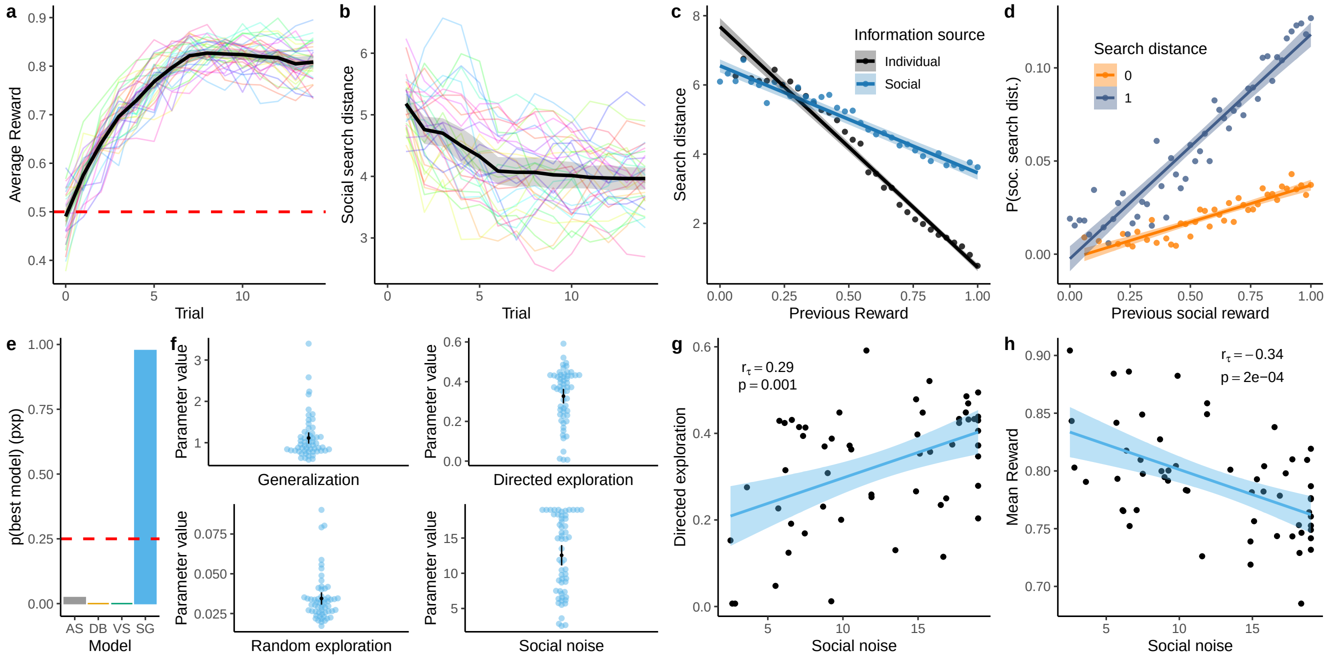

Figure 3. Experiment 1 results. a) Learning curves. The average reward across participants is shown in black, with group averages as coloured lines. The red dashed line shows chance-level performance. **b**) Social search distance (average Euclidian distance from other participants) over trials. The black line is the population average, while group averages as shown as coloured lines. **c**) Search distance as a function of previous reward, split by information source. Lines are the posterior prediction of a Bayesian hierarchical regression, while points are data averaged across 20 bins. **d**) Value-dependent social search distance. A search distance of 0 is imitation, while a search distance of 1 means the participant explored an adjacent tile to a social observation. **e**) Model comparison, showing the protected exceedance probability (pxp), which describes the probability of a model best fitting the population, accounting for chance. Red dashed line shows chance level. **f**) Social Generalization (SG) parameters (limited to participants best fit by SG). **g**) β (directed exploration) over ε_{soc}. Higher values of ε_{soc} mean lower reliance on social information. Only participants best fit by SG are shown. **h**) Mean reward over ε_{soc} (social noise). Higher values of ε_{soc} mean lower reliance on social information. Only participants best fit by SG are shown.

social) shows that participants' individual search distance (the Euclidian distance from their previous choice) significantly decreased as previous individual reward increased (-6.95; Highest Density Interval [-7.24, -6.67]; Fig. 3c, black line). This is rational given the spatial correlations in the environment, and is consistent with predictions of all candidate models. However, they did not only show this tendency for individual, but also for social information: when another participant earned a high reward in the previous trial, they searched closer to this participant's position (-4.22 [-4.61, -3.88]; Fig. 3c blue line). It is worth noting that this effect of social reward on search distance is significantly lower than the effect of individual information (2.73 [2.63, 2.79]), reflecting the lower reliability of social information compared to individual information.

This value-sensitive social information use is in line with predictions made by VS and SG, which integrate social information based on their value. On the other hand, it does not match predictions made by AS or DB: AS predicts no reliance on social information at all, with social search distances at roughly chance level (5.75), while DB would predict low social search distance regardless of previous social reward.

Further teasing apart the model predictions about social search distance, we consider the distinction of imitation (search distance = 0) vs. "innovation" (building on someone else's choice, search distance = 1; Fig. 3d). VS predicts value-sensitive imitation, that is, an increase of imitation rate as previous social reward increases, which we find in a linear model of social search distance frequency (0.04, 95%-CI: $[0.03, 0.05]$, $p < .001$). However, this effect was even stronger for innovation (0.08, 95%-CI: $[0.06, 0.10]$, $p < .001$), which only SG, the only model generalizing social information, could explain.

Modeling results

Turning to modelling, we find that Social Generalization did indeed fit human behaviour best, with hierarchical Bayesian model selection[46] showing it had the highest posterior probability of being the best model (protected exceedance probability: $pxp_{SG} \approx .98$; Fig. 3e). In participants best described by SG (Fig. 3f), the generalization parameter was significantly lower than the ground truth of $\lambda = 2$ ($\lambda \approx 1.11$; $t(56) = -13.0$, $p < .001$, $d = 1.7$, $BF > 100$). This means that participants did not generalize their observations as broadly as would be optimal given the environment. The directed exploration parameter was significantly lower than values found for individually learning GP-UCB agents in the same task structure[40] ($\beta \approx 0.33$; $t(56) = -9.4$, $p < .001$, $d = 1.3$, $BF > 100$). The random exploration parameter $\tau \approx 0.03$. Social noise was significantly higher than the value of 3.29 found to be optimal in evolutionary simulations ($\varepsilon_{soc} \approx 12.55$; $t(56) = 12.8$, $p < .001$, $d = 1.7$, $BF > 100$), meaning participants relied less on social information than optimal.

We find a relationship between β and ε_{soc}: the more a participant relied on social information (lower ε_{soc}), the less they relied on directed exploration (lower β; $r_\tau = .29$, $p = .001$, $BF = 28$; Fig. 3g). This might explain why β-values were lower than in previous, individual learning, settings[40]. Additionally, participants best described by SG performed better when they showed higher reliance on social information ($r_\tau = -.43$, $p < .001$, $BF > 100$; Fig. 3h), where the negative correlation reflects the fact that higher values of ε_{soc} mean lower reliance on social information.

In summary, Exp. 1 shows that participants could use social information to guide their decision-making even when it was not directly applicable to their own situation. Their behaviour followed the predictions of the SG model, implying that they used social information similarly to individual information, but treated it as more noisy, and thus less reliable. This method of integrating social information is optimal in our task environment (Fig. 2d), and lead to better results for participants the more they relied on social information. It is important to note that the linear relationship between social noise and reward only exists because participants relied on social information less than optimal. Indiscriminately using social information ($\varepsilon_{soc} = 0$) is not beneficial in our task (Fig. S2c), so the expected relationship when the whole range of ε_{soc} is covered would be U-shaped with lower rewards for both higher and lower reliance on social information than optimal. In addition, we find low uncertainty directed exploration (lower β) correlates with greater reliance on social learning (lower ε_{soc}). This pattern suggests social learning may partially replace uncertainty-directed exploration, which we expand on in Exp. 2.

Experiment 2. Social learning partially replaces directed exploration

To understand the effects of social information on decision-making better, we conducted a preregistered replication of Exp. 1 with the addition of solo rounds (i.e. rounds where participants were still in correlated environments, but were not shown other participants' choices and outcomes) to provide an asocial baseline for each participant. This allows us to control for the generic effects of the correlated reward structure, and directly probe if directed exploration was actually lower in social learning settings than in individual. This was a preregistered experiment[43]. Any analyses that were not included in the preregistration are specified as exploratory.

Behavioral results

Participants improved throughout trials, with higher performance on average in group rounds compared to solo rounds ($t(131) = 6.0$, $p < .001$, $d = 0.5$, $BF > 100$; Fig. 4a). Again, there was a minimal learning effect over rounds, but no effect of condition order or their interaction on performance (Fig. S1b).

In following analyses, we will compare social measures (like previous social reward, or social search distance) for both solo and group rounds despite no social information being provided in solo rounds. This serves to add a baseline for effects that could be interpreted as social (e.g. lower search distances for high previous social rewards) that might also be explained by participants independently exploring correlated environments. Social search distance decreased over trials in both conditions (Fig. 4b). However, it was significantly lower in group than in solo rounds ($t(131) = -14.8$, $p < .001$, $d = 1.7$, $BF > 100$), indicating that the clustering was not solely due to the social correlations between environments, but was influenced by social information.

Again, we investigate the effect of previous reward on search distance, splitting by information source (individual vs. social) and round type (solo vs. group). We replicate the results from Exp. 1 in group rounds: search distance was modulated by both individual (-7.75 [-8.00, -7.49]; Fig. 4c, black line) and social previous rewards (-5.44 [-5.75, -5.12]; Fig. 4c, dark blue line), with participants searching closer for higher values, and searching at greater distances for larger values. Again, social rewards influenced search distance to a lesser extent than individual rewards (2.31 [2.18, 2.44]). In solo rounds, we find the same effect for individual information (-7.97 [-8.22, -7.72]; Fig. 4c, gray line) with only a slight difference from group rounds (0.22 [0.02, 0.41]). This indicates that participants relied on previous individual reward slightly more in solo than in group rounds. Although previous social reward still significantly influenced social search distance in solo rounds, based on the correlated environmental structure only (-4.29 [-4.59, -4.00]; Fig. 4c, light blue line), it did so to a significantly lower degree than in group rounds (-1.15 [-1.33, -0.96]). This shows that, while the effect of social information on search distance can be partially explained by the socially correlated structure of the task, there is a significant component that can only be attributed to the use of social information in how participants modulated their search. Again, this result is in line with the predictions of VS and SG, but not Asocial Learning and Decision Biasing, which predict either no or indiscriminate social information use, respectively.

Focusing on a finer delineation of social search distance in an exploratory analysis (Fig. 4d), we find a value-based increase in imitation frequency (0.023, 95% CI: $[0.011, 0.035]$, $p < .001$) and even higher increase in innovation (0.062, 95%-CI: $[0.046, 0.078]$, $p < .001$) across round types. However, this increase in frequency was also significantly higher in group rounds compared to solo rounds (0.036, 95%CI:

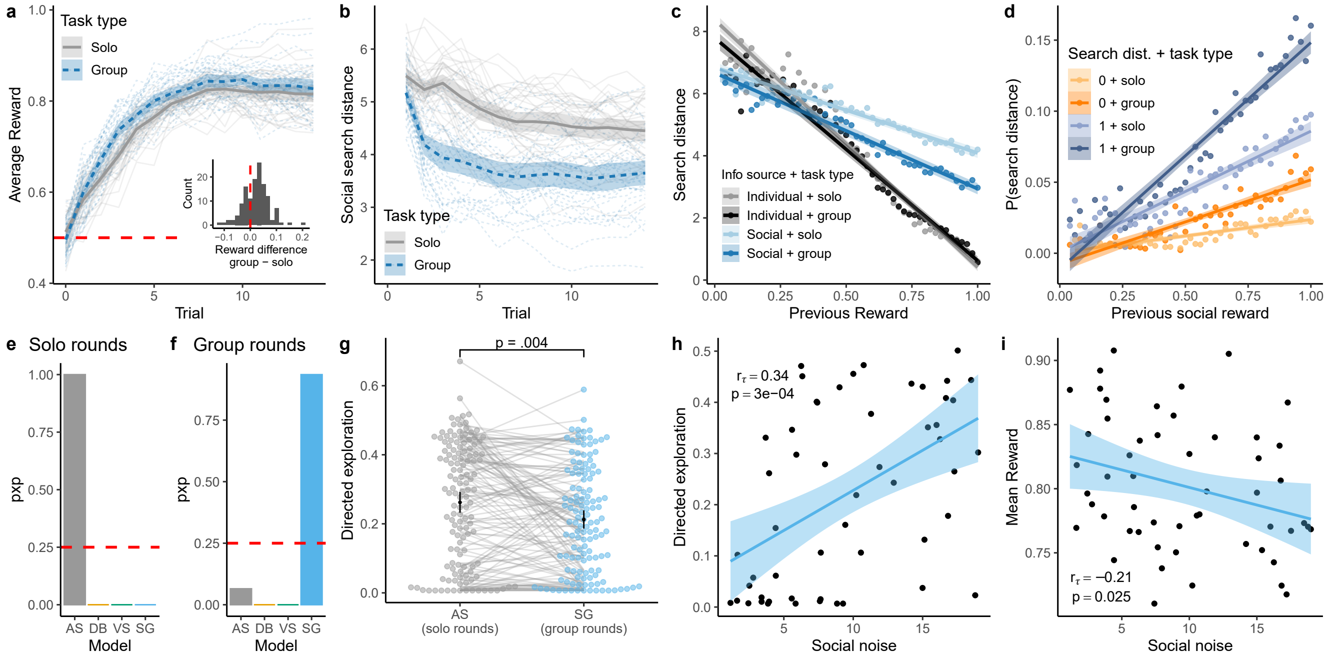

Figure 4. Exp. 2 results. **a**) Learning curves by task type. Averages for solo (gray solid) and group (blue dashed) rounds shown in thick lines with shaded 95% CIs, while group averages are shown as thin lines. Red dashed line shows chance-level performance. Inset plot is an histogram of mean score in group - solo rounds, with the red dashed line showing no difference. **b**) Social search distance over trials by task type. Averages for solo (gray solid) and group (blue dashed) rounds thick and shaded, group level averages thin lines. **c**) Search distance over previous reward split by information source and task type. Lines are results of Bayesian hierarchical regression fixed effects, points are data averaged across 20 bins. Solo rounds in pale and group rounds in saturated colours, individual information in gray and social information in blue. **d**) Value dependent social search distance by task type. Solo rounds are solid lines, group rounds dashed. **e-f**) Protected exceedance probabilities (pxps) for solo (e) and group (f) rounds. Probability of a model best fitting the population, accounting for chance. Red dashed line shows chance level. **g**) β-parameter estimates for participants fit by AS in solo rounds (gray) and participants fit by SG in group rounds (blue). For all SG parameters, see Fig. S9. **h**) β (directed exploration) over ε_{soc} in group rounds. Higher values of ε_{soc} mean lower reliance on social information. Only participants best fit by SG are shown. **i**) Mean reward over ε_{soc} in group rounds. Higher values of ε_{soc} mean lower reliance on social information. Only participants best fit by SG shown.

$[0.019, 0.053]$, $p < .001$), and higher still for innovation in group rounds (0.038, 95%CI: $[0.016, 0.061]$, $p = .001$). Thus, we replicate the value-sensitive increase in both imitation and innovation, which is only predicted by SG. The significant interaction of the effect with round type once again shows that, while the effects found in Exp. 1 can be partially explained by the correlation structure of the environments alone, they remain significant when controlling for this factor.

Modeling results

We again performed hierarchical Bayesian model comparisons, but separately for solo and group rounds. Here, we find that AS is the best fitting model for solo rounds, showing that our social models did not just exploit the correlated structure to improve fit (Fig. 4e, $pxp \approx 1$). In group rounds, we again find that SG is the best fitting model (Fig. 4f; $pxp \approx .94$).

In line with our finding of comparatively lower values of directed exploration parameter β in Exp. 1 than in previous individual learning literature, we find that β is indeed significantly lower in solo than in group rounds within participants (Wilcoxon signed-rank test; $Z = -2.7$, $p = .004$, $r = -.23$, $BF = 63$; Fig. 4g). In an exploratory analysis, we replicate the significant relationship between ε_{soc} and β from Exp. 1 ($r_\tau = .34$, $p < .001$, $BF > 100$; Fig. 4h), again suggesting a partial replacement of directed exploration with social learning when social information is available.

Regarding the relationship of social noise parameter ε_{soc} and average reward in group rounds in participants best fit by SG, we find a weakly significant correlation ($r_\tau = -.21$, $p = .025$, $BF = 2.1$, Fig. 4i). This might be explained by a ceiling effect of social learning that was not as strong in Exp. 1, when they had fewer rounds to familiarize themselves with the task.

In summary, Exp. 2 replicates the findings of experiment 1 in that participants use social information even when it is not directly applicable to their own situation, being best described by the SG model. This means that social information is used similarly to individual information, but treated as more noisy.

The addition of an asocial baseline condition lets us compare individual and social strategies, where we confirm that the findings in Exp. 1 were actually indicative of social information use and not just a consequence of purely individual information use in a correlated environment. We replicate the finding that β-values in group rounds are lower than in previous literature. In comparison to solo rounds, we find that β-values are significantly lower, and significantly correlated with social noise, indicating that the use of social information replaces uncertainty-directed exploration to a degree. Use of social information was beneficial, shown both by the negative correlation of social noise and mean reward, and the general higher scores in group compared to solo rounds.

Discussion

We introduce the socially correlated bandit as a task to study social learning with similar, but not identical reward structures. Here, we find that the best normative and descriptive model, Social Generalization (SG), integrates social information in a noisy fashion, thus extending past research on social learning in settings using the same task environment for the source of social information and the participant[1,2,4,42].

The socially correlated bandit has a spatially correlated reward structure within participant, which is also positively correlated across participants. The spatial correlations allows for social information to be integrated at a stage of the decision-making process unique to tasks in which reward information can be generalized. Thus, we introduce the SG model, which generalizes social information similarly to individual information, but treats it as more noisy (Eq. 7). We also show that the previously dominant Value Shaping (VS) model[2] can be seen as an edge case of SG in cases where participants are in identical environments and fully relying on social information as would be sensible when learning from expert demonstrators (Fig. S4).

In Exp. 1, we found that SG was the best descriptive model of human behaviour in this task, which was additionally confirmed by behavioural patterns that were only consistent with this model's predictions about the integration of social information into participants' search behaviour (Fig. 3b-d). We also found a relationship between the social noise parameter ε_{soc} and performance, implying that participants were more successful the more they relied on social information (Fig. 3h). Additionally, we found that participants who relied more on social information displayed less uncertainty-directed exploration (Fig. 3g). This suggests a potential replacement of exploration with social learning, further corroborated by lower values of the directed exploration parameter β than found in previous works using an individual version of the task[40,44], which motivated further investigation in Exp. 2.

In Exp. 2, we conducted a preregistered replication[43] of Exp. 1, which also added solo rounds to the task to assess how exploration behaviour changed within participants when social information is available. The within-subject manipulation of solo vs. group rounds allowed us to ensure that none of the findings of Exp. 1 were solely the consequence of individual learning in correlated environments, and let us compare patterns and outcomes of individual and social learning. In this experiment, participants performed better in group than solo rounds, showing that they used available social information to their benefit (Fig. 4a). Again, SG was the best descriptive model of human behaviour (Fig. 4f). We replicate the behavioural signatures corroborating SG as the winning model even while accounting for the individual learning baseline (Fig. 4b-d). We indeed found that β was significantly lower in group than in solo round, with β again being significantly correlated with social noise ε_{soc} (Fig. 4h), further lending credibility to social learning replacing directed individual exploration.

Across both experiments, we expand task settings for computational models of social information integration to cases where imitation is not optimal, and find that humans can do more than imitate when the situation calls for it[10]. Our task expands the niche case of group settings with identical reward environments to non-identical, but positively correlated environments in line with how in real life many situations in which we learn from others call for some distinction between the other person's situation and ourselves. Previous research can be mapped as a fringe case of our setup, and SG can be seen as an adaptation of a previously winning[2] and, based on its capacity for both memory and value-sensitivity, most flexible model, Value Shaping, to settings where agents are not learning from expert demonstrators (Fig. S4).

In sum, the findings of our study add to a rich literature showing that social learning is adaptive in stable environments[1,42,47–52] even in non-identical environments. We show that this adaptive use of social information went hand in hand with a reduction of uncertainty-directed exploration, implying that social learning functioned as an exploration-tool.

Social learning and resource rationality

While we find adaptive use of social information for our task setting, at the same time, we still find that participants underutilized social information compared to what would be optimal, as has been previously found in experimental settings[5,24,48,53,54]. Human's natural skill at social learning may be limited by the artificial experimental setting. A part of this may be that some social learning strategies, like copying the expert, were impossible due to lack of information, potentially reducing participants' inclination to rely on social information overall.

However, besides social learning potentially being impeded by the artificial experimental setting, the discrepancy between adaptive social learning and underutilization of social information may also be explained by resource rationality. While it may be theoretically optimal to discount social information to a specific, low degree, this may also be significantly more complex than to rely on individual information more strongly, only referring back to presumably noisy social information when individual learning does not provide any promising options. In this regard, underutilizing social information may also be seen as resource-rational in that regard[55]. Falling back on social information only when it is absolutely necessary also ties back to the social learning strategies[8,10]: Social Generalization agents generally copy (i.e. are strongly influenced by others' choices) when uncertain (i.e. their individual information does not outweigh social observations).

The same resource-rationality based reasoning may be applied to our finding regarding directed exploration being partially replaced by social learning when possible. Directed exploration has been shown to be reduced by cognitive load[56,57], indicating that individual exploration may be costly. Hence, our finding that social learning may have served as an exploration tool in our task hints at social learning as a method to let us offload these costs of directed exploration. It is also in line with prior research that suggests or shows exploration differences between asocial and social settings[1,27,41,42,49,58,59]. It also provides empirical support for the outcome of the social learning strategies tournament, wherein winning models tended to almost exclusively use social learning for exploration[9].

In our task, we show that this lower exploration is optimal for social learning using evolutionary simulations. However, simulations based on our participants' parameter estimates show that lower exploration would also lead to higher performance in asocial learning (Fig. S10a-b). This implies that social learning not only takes the function of uncertainty-directed exploration, but also helps participants avoid overexploration. This might be due to the need for exploration being diminished by the ability to gain more environmental information from others. It could also be a dynamic process wherein observing one teammate move from exploring to exploiting inspires participants to do the same, lowering overall exploration rates compared to individual settings. Such adjustments of strategy between individual and group settings, especially for individuals with low confidence, have been found before[47]. However, the exact mechanism of this lowered exploration in a round-based task remains a subject for further research. In an exploratory analysis, we find a similar effect of social learning on generalization parameter λ being higher, and thus closer to the ground truth, in group rounds (Fig. S10c-d), which is in line with previous research[42].

Limitations and future directions

Previous research often investigates the effects of demonstrator skill, contrasting one skilled and one unskilled demonstrator[2,60]. Following this reasoning, one might consider that the integration mechanism used depends on the skill level of teammates. However, value-sensitive models VS and SG benefit from any information about the structure of the environment, regardless of if it is positive or negative. Therefore, we did not investigate effects of participant skill directly. However, participants appeared more sensitive to choices of their peers that lead to high rewards (Fig. 3 and Fig. 4c), mirroring the human tendency to "copy" (here rather "learn from") the successful[24,25]. Nevertheless, it remains an open question how sensitive participants can be to others' perceived skill in this task, and in what way this would influence their decision-making.

With the focus of this study being on the mechanisms of integrating social information as an individual, we left group dynamics of exploration throughout the experiment largely unexplored. While we find social clustering on a group level (Fig. 3b and Fig. 4b), we can explain this using individual-level mechanisms. In our task design, participants are incentivized to maximize individual gain, which limits the benefit of active coordination. Based on participant's reported strategies, it is unlikely that they coordinated their behaviour to maximize information gain as well. However, it may be interesting to investigate coordination strategies in similar task settings, for example by changing the incentive[61] to optimal understanding of the environment rather than maximizing rewards.

Our experiments limited the environmental correlation to $r = 0.6$. This is due to the fact that it was both harder to generate many environments with higher correlations, and the results would be less insightful, likely converging on imitation as they approach 1. However, our task using only one specific correlation of environments leads to a number of new questions: For which range of correlations humans are still sensitive to the optimal strategy? When (if ever) do they stop integrating social information altogether? Are humans able to make use of negatively correlated environments as well as positively correlated ones? Additionally, in real life, we would expect to find some people with more similar tastes to us and others with more different tastes. How such varying correlations between participants would affect how social information is used remains an open question. Given prior research showing that humans are quite capable of adjusting their social learning based on the skill of the observed individual[2,25,26], it seems reasonable to assume they could adjust to higher or lower levels of correlation as well. It would be interesting to see if this would lead to only learning from the most closely correlated individual, or from all sources but with higher assumed noise for lower correlations.

Given the novel task setting of social learning in positively correlated environments, we chose to investigate the naturalistic interactions of groups of four real participants. We would not have been able to make an informed choice of model for a more controlled setting where humans are placed in groups of artificial agents a priori. Thus, having humans do the task in groups ensured that we were not affecting their behaviour through unnatural model choices. In the future, more granular insights into the exact usage of social information could be gained by placing participants in groups with Social Generalization agents, which can used to more precisely manipulate the usefulness of social information and control group dynamics.

Conclusions

Across two experiments, we found that people used social information more flexibly than previously accounted for, successfully integrating information from others with diverse reward functions, but taking it "with a grain of salt". Our model captures this, by integrating social information as inherently noisier, since it is not directly applicable to one's own circumstances. Social learning also functioned as an exploration tool, partially replacing uncertainty-directed exploration, and potentially helping participants behave more optimally.

Methods

Experiment design

Across both experiments, participants explored spatially correlated multi-armed bandits[40] with social correlations across participants. The bandits were displayed as grids consisting of 121 tiles. Environments were structured identically across studies and conditions. Each tile yielded normally distributed rewards: $r(\mathbf{x}) \sim \mathcal{N}(f(\mathbf{x}), \sigma_\varepsilon^2)$ where the expected reward across all tiles was sampled from a Gaussian Process (GP) prior to induce spatial correlations $f \sim \mathcal{GP}\left(0, k(\mathbf{x}, \mathbf{x}')\right)$ and the variance was fixed to $\sigma_\varepsilon^2 = .0001$. To generate the environments, we sampled a set of 11x11 parent grids from a Gaussian process prior with an RBF-kernel with a length scale of $\lambda = 2$. We then used these parent environment's means as the prior means to sample a set of child environments. To facilitate correlations across environments, the child environments were filtered to only include those which correlated with the parent environment by at least $r = .6$. This subset of child environments was then filtered to only include sets of 4 environments which had correlation coefficients of $r = .6 \pm 0.05$ with each other to use in the task. We generated 40 sets of correlated environments for the experiments this way. At the start of the experiment, we sampled a number of environment sets corresponding to the number of rounds without replacement for each group.

For each round of the experiment, each participant was assigned an environment from such a correlated set of four. Exp. 1 consisted of 8 rounds, and Exp. 2 consisted of 8 solo and 8 group rounds, totalling 16 rounds. The search horizon was 14 for both experiments, with one tile being revealed at random at the start of each round. To prevent participants from getting used to the same reward structure, including its global maximum, environments were rescaled to a randomly selected maximum value between 60 and 80 for each round. This rescaling was consistent across participants. To prevent a single participant from holding up a group, a random tile would be selected if they did not make a choice within 10 seconds. Such random choice trials were excluded from analysis. After selecting a tile, they would wait for all other participants to make their selection as well. Once all participants made a choice, the task would move on to the next trial. In *solo rounds*, participants would only see their own bandit. In *group rounds*, participants were also permanently shown all other participants' bandits, including choices and outcomes. This was the only difference between the two conditions. Choice and outcome information in group rounds were updated for all participants at once after all group members had made a choice.

Participants and design

Participants for both experiments were recruited via Prolific and assigned to groups of four based on access time to the experiment. They were paid a base rate for expected experiment duration, and could earn a bonus of maximum the same amount based on performance. Both experiments were approved by the Ethics Committee of the University of Konstanz ("Collective learning and decision-making study"), and participants provided informed consent prior to participation.

Exp. 1 was an observational study with only the group condition, for which we recruited *N=188* participants. After eliminating all groups with drop-out, the final sample size was *N=128* (mean age: 38.5 ± 12.7 SD; 44 females). On average, participants spent 20.8 ± 0.5 minutes on the task and earned £ 7.19 ± 0.04.

For Exp. 2, which varied solo vs. group conditions within-subject in interleaved order, we recruited 220 participants. Condition order (solo round first vs. group round first) was counterbalanced across groups. After eliminating all groups with drop-out, the final sample size was N=132 (mean age: $35.8 \pm 11.2 SD$, 46 females). On average, participants spent 31.0 ± 0.6 minutes on the task and earned £10.4 ± 0.08.

Materials and procedure

In both experiments, participants took part in groups of four, which they were assigned to based on access time. After giving informed consent, participants were instructed that they were embarking on a scientific mission to collect salt samples from alien oceans on other planets, and that their goal was to collect as many salt samples as possible. They were informed that they could revisit the same area to get a similar reward, with salt not depleting from repeated sampling. They were also told that other scientists on their team would collect different salts, so there would be no competition for resources, but that the salts were generated by the same process, and locations with high salt concentrations were thus correlated across the salts. In Exp. 2, participants were additionally told that they would be sent on both solo and group missions, with no information from their teammates being available in solo missions. Participants in both experiments were shown fully revealed example environments to ensure they understand the structure and how social information usage may benefit them. After passing a comprehension check, participants moved on to a waiting room, which would launch the task once four people had joined. If there was no group of four after 3 minutes of a room being open, all participants in that room were redirected to the post experiment questionnaire.

Once in the task, participants would be presented with their bandit grid with one tile revealed. In the group condition, they would additionally see the bandits of all other participants in the group with the rewards revealed as well. While participants' bandits were still correlated in solo rounds, they could not see other group members' bandits or choices in this condition.

Evolutionary simulations

For any possible combination of models, we generated an initial population consisting of an equal proportion of all the models. For the three-way mixes, where have exactly equal numbers, the final agent was randomly selected to be any of the three models. Initial populations were generated based on a common set of priors (Fig. 2b, see also SI). We used tournament selection to select agents for the next generation: groups of four agents were randomly drawn with replacement to compete in one round of the task. The selection probability of agents thus selected was lowered to prevent the same agents from being sampled too often. The agent with the highest score in a group was selected to seed the next generation. This procedure was repeated until the full population size of N=100 was reached. Each agent was thus sampled about 4 times. Before repeating the process for the next generation, mutations were applied to a part of the population. There was a 2% chance of *parameter mutations*, in which a parameter would have Gaussian noise $\sim \mathcal{N}(0, 0.2)$ added. If this caused the parameter to go out of bounds, it was resampled from prior. There was a 0.2% chance of a *type mutation*, in which the agent's model would be randomly resampled. The new model could be one that was not initially present in the population. To allow for invasions, we kept the baseline (GP-UCB) parameters of the mutating agent stable, and only modified the social parameter, which determines the model. Simulations were run this way for 500 generations. Simulations of all initial populations were repeated 10 times to ensure stability of the results.

Model comparisons

We fit models based on cross-validated maximum-likelihood estimation. We iteratively formed the training sets by leaving one round out, computing parameter estimates on this set, and evaluating model predictions on the out-of-sample round. Overall goodness of fit was evaluated based on the sum of the prediction error on each of the out-of-sample predictions. For Exp. 2, participant data was split into solo and group rounds before fitting. We used the summed out-of-sample log likelihood as an approximation of the model evidence to perform hierarchical Bayesian model comparison[46].

References

1. Toyokawa, W., Whalen, A. & Laland, K. N. Social learning strategies regulate the wisdom and madness of interactive crowds. *Nat. Hum. Behav.* **3**, 183–193 (2019).
2. Najar, A., Bonnet, E., Bahrami, B. & Palminteri, S. The actions of others act as a pseudo-reward to drive imitation in the context of social reinforcement learning. *PLoS biology* **18**, e3001028 (2020).
3. Biele, G., Rieskamp, J., Krugel, L. K. & Heekeren, H. R. The neural basis of following advice. *PLoS biology* **9**, e1001089 (2011).
4. Park, S. A., Goïame, S., O'Connor, D. A. & Dreher, J.-C. Integration of individual and social information for decision-making in groups of different sizes. *PLoS biology* **15**, e2001958 (2017).
5. Molleman, L. *et al.* Strategies for integrating disparate social information. *Proc. Royal Soc. B* **287**, 20202413 (2020).
6. Charpentier, C. J., Iigaya, K. & O'Doherty, J. P. A neuro-computational account of arbitration between choice imitation and goal emulation during human observational learning. *Neuron* **106**, 687–699 (2020).
7. Rendell, L. *et al.* Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends cognitive sciences* **15**, 68–76 (2011).
8. Laland, K. N. Social learning strategies. *Animal Learn. & Behav.* **32**, 4–14 (2004).
9. Rendell, L. *et al.* Why copy others? Insights from the social learning strategies tournament. *Science* **328**, 208–213 (2010).
10. Kendal, R. L. *et al.* Social learning strategies: Bridge-building between fields. *Trends Cogn. Sci.* **22**, 651–665 (2018).
11. Hoppitt, W. & Laland, K. N. *Social learning: an introduction to mechanisms, methods, and models* (Princeton University Press, 2013).
12. Toyokawa, W., Saito, Y. & Kameda, T. Individual differences in learning behaviours in humans: Asocial exploration tendency does not predict reliance on social learning. *Evol. Hum. Behav.* **38**, 325–333 (2017).
13. Deffner, D., Kleinow, V. & McElreath, R. Dynamic social learning in temporally and spatially variable environments. *Royal Soc. open science* **7**, 200734 (2020).
14. Morgan, T. J., Rendell, L. E., Ehn, M., Hoppitt, W. & Laland, K. N. The evolutionary basis of human social learning. *Proc. Royal Soc. B: Biol. Sci.* **279**, 653–662 (2012).
15. Kameda, T. & Nakanishi, D. Cost–benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evol. Hum. Behav.* **23**, 373–393 (2002).
16. Heath, C., Bell, C. & Sternberg, E. Emotional selection in memes: the case of urban legends. *J. personality social psychology* **81**, 1028 (2001).
17. Eriksson, K. & Coultas, J. C. Corpses, maggots, poodles and rats: Emotional selection operating in three phases

of cultural transmission of urban legends. *J. Cogn. Cult.* **14**, 1–26 (2014).

18. Stubbersfield, J. M., Tehrani, J. J. & Flynn, E. G. Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends. *Br. journal psychology* **106**, 288–307 (2015).

19. Blaine, T. & Boyer, P. Origins of sinister rumors: A preference for threat-related material in the supply and demand of information. *Evol. Hum. Behav.* **39**, 67–75 (2018).

20. Stubbersfield, J. M., Tehrani, J. J. & Flynn, E. G. Chicken tumours and a fishy revenge: Evidence for emotional content bias in the cumulative recall of urban legends. *J. Cogn. Cult.* **17**, 12–26 (2017).

21. Mesoudi, A., Whiten, A. & Dunbar, R. A bias for social information in human cultural transmission. *Br. journal psychology* **97**, 405–423 (2006).

22. Brand, C., Heap, S., Morgan, T. & Mesoudi, A. The emergence and adaptive use of prestige in an online social learning task. *Sci. reports* **10**, 1–11 (2020).

23. Brand, C., Mesoudi, A. & Morgan, T. J. Trusting the experts: The domain-specificity of prestige-biased social learning. *PloS one* **16**, e0255346 (2021).

24. Mesoudi, A. An experimental comparison of human social learning strategies: payoff-biased social learning is adaptive but underused. *Evol. Hum. Behav.* **32**, 334–342 (2011).

25. Watson, R., Morgan, T. J., Kendal, R. L., Van de Vyver, J. & Kendal, J. Social learning strategies and cooperative behaviour: Evidence of payoff bias, but not prestige or conformity, in a social dilemma game. *Games* **12**, 89 (2021).

26. Wu, C. M. *et al.* Visual-spatial dynamics drive adaptive social learning in immersive environments. *bioRxiv* DOI: 10.1101/2023.06.28.546887 (2023).

27. Toyokawa, W. & Gaissmaier, W. Conformist social learning leads to self-organised prevention against adverse bias in risky decision making. *Elife* **11**, e75308 (2022).

28. Galef, B. G. Imitation in animals: history, definition, and interpretation of data from the psychological laboratory. In *Social learning*, 15–40 (Psychology Press, 2013).

29. Analytis, P. P., Barkoczi, D. & Herzog, S. M. Social learning strategies for matters of taste. *Nat. human behaviour* **2**, 415–424 (2018).

30. Müller-Trede, J., Choshen-Hillel, S., Barneron, M. & Yaniv, I. The wisdom of crowds in matters of taste. *Manag. Sci.* **64**, 1779–1803 (2018).

31. Yaniv, I., Choshen-Hillel, S. & Milyavsky, M. Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organ. Behav. Hum. Decis. Process.* **115**, 111–120 (2011).

32. Jara-Ettinger, J. Theory of mind as inverse reinforcement learning. *Curr. Opin. Behav. Sci.* **29**, 105–110 (2019).

33. Jara-Ettinger, J., Schulz, L. E. & Tenenbaum, J. B. The naive utility calculus as a unified, quantitative framework for action understanding. *Cogn. Psychol.* **123**, 101334 (2020).

34. Shafto, P., Goodman, N. D. & Frank, M. C. Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychol. Sci.* **7**, 341–351 (2012).

35. Croom, S., Zhou, H. & Firestone, C. Seeing and understanding epistemic actions. *Proc. Natl. Acad. Sci.* **120**, e2303162120 (2023).

36. Berke, M. & Jara-Ettinger, J. Integrating experience into bayesian theory of mind. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 44 (2022).

37. Hawkins, R. D. *et al.* Flexible social inference facilitates targeted social learning when rewards are not observable. *Nat. Hum. Behav.* **7**, 1767–1776 (2023).

38. Baker, C. L., Jara-Ettinger, J., Saxe, R. & Tenenbaum, J. B. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* **1**, 0064 (2017).

39. Jern, A., Lucas, C. G. & Kemp, C. People learn other people's preferences through inverse decision-making. *Cognition* **168**, 46–64 (2017).

40. Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D. & Meder, B. Generalization guides human exploration in vast decision spaces. *Nat. human behaviour* **2**, 915–924 (2018).

41. Plate, R. C., Ham, H. & Jenkins, A. C. When uncertainty in social contexts increases exploration and decreases obtained rewards. *J. Exp. Psychol. Gen.* (2023).

42. Naito, A., Katahira, K. & Kameda, T. Insights about the common generative rule underlying an information foraging task can be facilitated via collective search. *Sci. Reports* **12**, 1–12 (2022).

43. Witt, A., Toyokawa, W., Gaissmaier, W., Lala, P., Kevin N & Wu, C. M. Social learning in a spatially correlated multi- armed bandit, DOI: 10.17605/OSF.IO/DVH6Y (2024).

44. Giron, A. P. *et al.* Developmental changes in exploration resemble stochastic optimization. *Nat. Hum. Behav.* DOI: 10.1038/s41562-023-01662-1 (2023).

45. Rogers, A. R. Does biology constrain culture? *Am. Anthropol.* **90**, 819–831 (1988).

46. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies—revisited. *Neuroimage* **84**, 971–985 (2014).

47. Tump, A. N., Pleskac, T. J. & Kurvers, R. H. Wise or mad crowds? The cognitive mechanisms underlying information cascades. *Science Advances* **6**, eabb0266 (2020).

48. Tump, A. N., Wolf, M., Krause, J. & Kurvers, R. H. Individuals fail to reap the collective benefits of diversity because of over-reliance on personal information. *J. Royal Soc. Interface* **15**, 20180155 (2018).

49. Toyokawa, W., Kim, H.-r. & Kameda, T. Human collective intelligence under dual exploration-exploitation dilemmas. *PloS one* **9**, e95789 (2014).

50. Wolf, M., Krause, J., Carney, P. A., Bogart, A. & Kurvers, R. H. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PloS one* **10**, e0134269 (2015).

51. Bang, D. & Frith, C. D. Making better decisions in groups. *Royal Soc. open science* **4**, 170193 (2017).

52. Bahrami, B. *et al.* Optimally interacting minds. *Science* **329**, 1081–1085 (2010).

53. Morin, O., Jacquet, P. O., Vaesen, K. & Acerbi, A. Social information use and social information waste. *Philos. Transactions Royal Soc. B* **376**, 20200052 (2021).

54. Acerbi, A., Tennie, C. & Mesoudi, A. Social learning solves the problem of narrow-peaked search landscapes: experimental evidence in humans. *Royal Soc. open science* **3**, 160215 (2016).

55. Bhui, R., Lai, L. & Gershman, S. J. Resource-rational decision making. *Curr. Opin. Behav. Sci.* **41**, 15–21 (2021).

56. Cogliati Dezza, I., Cleeremans, A. & Alexander, W. Should we control? the interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *J. Exp. Psychol. Gen.* **148**, 977 (2019).

57. Wu, C. M., Schulz, E., Pleskac, T. J. & Speekenbrink, M. Time pressure changes how people explore and respond to uncertainty. *Sci. reports* **12**, 4122 (2022).

58. Cohen, J. D., McClure, S. M. & Yu, A. J. Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philos. Transactions Royal Soc. B: Biol. Sci.* **362**, 933–942 (2007).

59. Fleischhut, N., Artinger, F. M., Olschewski, S. & Hertwig, R. Not all uncertainty is treated equally: Information search under social and nonsocial uncertainty. *J. Behav. Decis. Mak.* **35**, e2250 (2022).

60. Selbing, I., Lindström, B. & Olsson, A. Demonstrator skill modulates observational aversive learning. *Cognition* **133**, 128–139 (2014).

61. Deffner, D. *et al.* Collective incentives reduce over-exploitation of social information in unconstrained human groups, DOI: 10.31234/osf.io/p3bj7 (2023).

62. Witt, A., Toyokawa, W., Lala, K., Gaissmaier, W. & Wu, C. M. Social learning with a grain of salt. In Goldwater, M., Anggoro, F., Hayes, B. & Ong, D. (eds.) *Proceedings of the 45th Annual Conference of the Cognitive Science Society*, DOI: 10.31234/osf.io/c3fuq (Cognitive Science Society, Sydney, Australia, 2023).

Data and Code Availability

All data and code will be made publicly available upon publication.

Acknowledgements

AW and CMW are supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting AW. WT and WG are supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2117 – 42203798. An early version of Exp. 1 was presented at the 45th Annual Conference of the Cognitive Science Society[62].

Author contributions statement

CMW, WT, and AW conceived the experiments. WT and AW conducted the experiments. AW and CMW analysed the results and wrote the manuscript. All authors reviewed the manuscript.

Supplementary Information for

Flexible integration of social information despite interindividual differences in reward.

Alexandra Witt, Wataru Toyokawa, Kevin N. Lala, Wolfgang Gaissmaier & Charley M. Wu

Learning and ordering effects

In Exp. 1, there was a small learning effect over rounds ($0.004, 95\% - CI: [0.001, 0.007], p = 0.0027$). The same was true for Exp. 2 ($0.002, 95\% - CI: [0.0003, 0.003], p = 0.012$). In Exp. 2, there was no effect of block order ($0.005, 95\% - CI: [-0.02, 0.015], p = 0.63$), or the interaction of round and block order ($-0.0001, 95\% - CI: [-0.002, 0.002], p = 0.94$) on performance (Fig. S1).

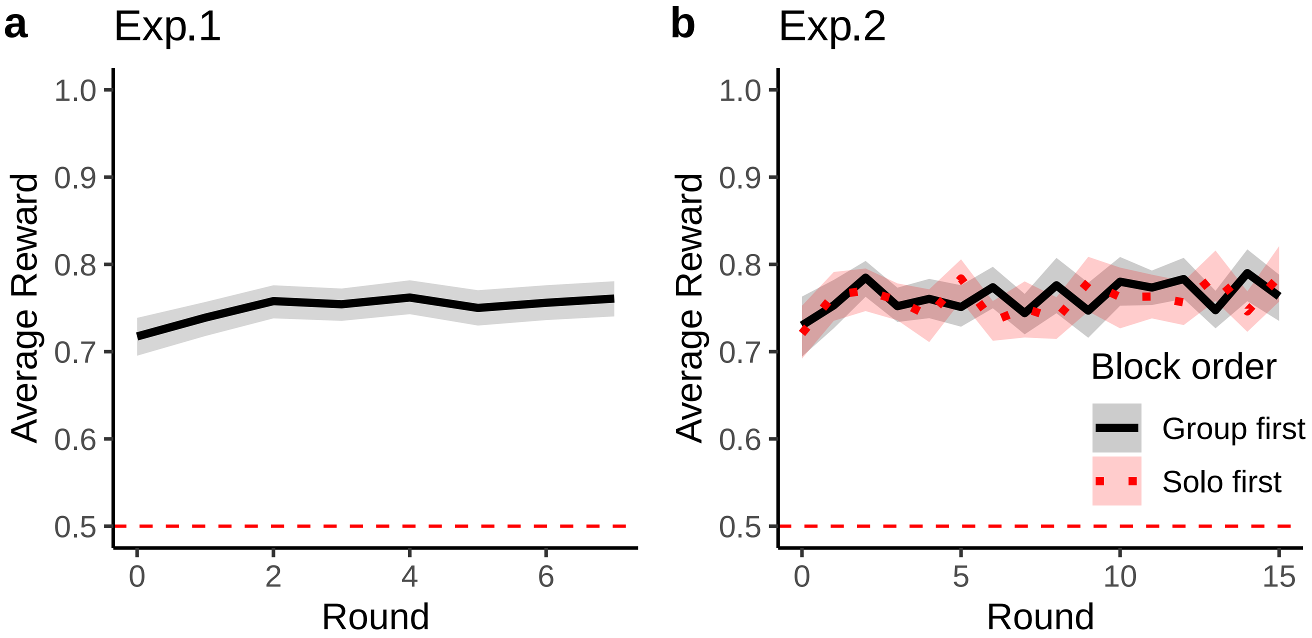

Figure S1. Learning over rounds. a) Learning over rounds in Exp. 1. Red dashed line gives chance level performance. **b**) Learning over rounds and by block order in Exp. 2. Red dashed line gives chance level performance.

Model simulations and variants

Priors

For initial agent-based and evolutionary simulations, we drew parameters from a set of parameter priors. Priors for the baseline asocial learner were based on the values found in prior studies[40]. Since none of the parameter values can be negative, but have no upper bound, we used log-normal distributions around the reported average participant estimates. This resulted in the following parameter priors:

$$\begin{aligned} \lambda, \beta &\sim \text{LogNormal}(0.75, 0.5) \\ \tau &\sim \text{LogNormal}(4.5, 0.75) \end{aligned} \tag{8}$$

For the social parameters, no prior empirical results existed, so we used priors that covered as much of the theoretical space as possible. While we are able to cover the entire possible range for Decision Biasing and Value Shaping, since $\alpha, \gamma \in [0, 1]$, the Social Generalization noise parameter ε_{soc} cannot be negative, but can grow infinitely large. Therefore, we chose to centre an exponential distribution around $\varepsilon_{soc} = 2$, which we found to be good, but not optimal, in simulations, resulting in the following priors:

$$\begin{aligned} \alpha, \gamma &\sim \text{Uniform}(0,1) \\ \varepsilon_{soc} &\sim \text{Exponential}(0.5) \end{aligned} \tag{9}$$

Unless otherwise stated, simulations were run using these priors.

Model variants

For model variants, we simulated groups of two asocial agents as well as one canonical and one modified agent. This was done to be able to directly compare the model's performance given identical information. Asocial agents were used to prevent Roger's paradox[45], i.e. the frequency-dependent fitness of social models, from affecting the results. We ran these simulations with task settings (search horizon of 14, 8 rounds) with 1000 different parameter sets.

Value-agnostic Value Shaping

As mentioned in the main text, Value Shaping benefitted from including social value information (Fig. S2a) compared to the unbiased version common in previous literature, where value is generically boosted for options selected by others. We implemented this as $V(\mathbf{x}) = V_{x,ind} + \alpha \cdot n_{x_{soc},t-1}$. Our canonical prediction error implementation of VS significantly outperforms this alternative ($t(1998) = 8.4$, $p < .001$, $d = 0.4$, $BF > 100$) and was thus chosen as the main implementation.

Value-sensitive Decision Biasing

We also tried to adapt Decision Biasing to our task by including value information. This might have been a beneficial update, since outcome information was generally not available in previous studies[1,2], but was in ours. We modified the social policy so that it increased proportionally to the frequency of a social choice, weighted by how much higher the social reward was than the average experienced individual reward $\pi_{soc}(x) \propto n_{x_{soc_{t-1}}} \cdot (m(x_{soc_{t-1}} - m(\bar{x}_{ind})$. In cases where $m(x_{soc_{t-1}} < m(\bar{x}_{ind})$, the social information was ignored to prevent negative probabilities in the policy. Despite this added information (and the large sample size), there was no significant difference between the two models' scores ($t(1998) = 0.3$, $p = .730$, $d = 0.02$, $BF = .05$; Fig. S2a). Thus, we chose to use the simpler model.

Indiscriminate Social Generalization

There is an edge case of Social Generalization for $\varepsilon_{soc} = 0$. It means that social and individual information are treated identically. While technically not a separate model, we show that discriminate use of social information ($\varepsilon_{soc} \neq 0$ is significantly better than indiscriminate use in our task ($t(1998) = 3.4$, $p < .001$, $d = 0.2$, $BF = 18$; Fig. S2b).

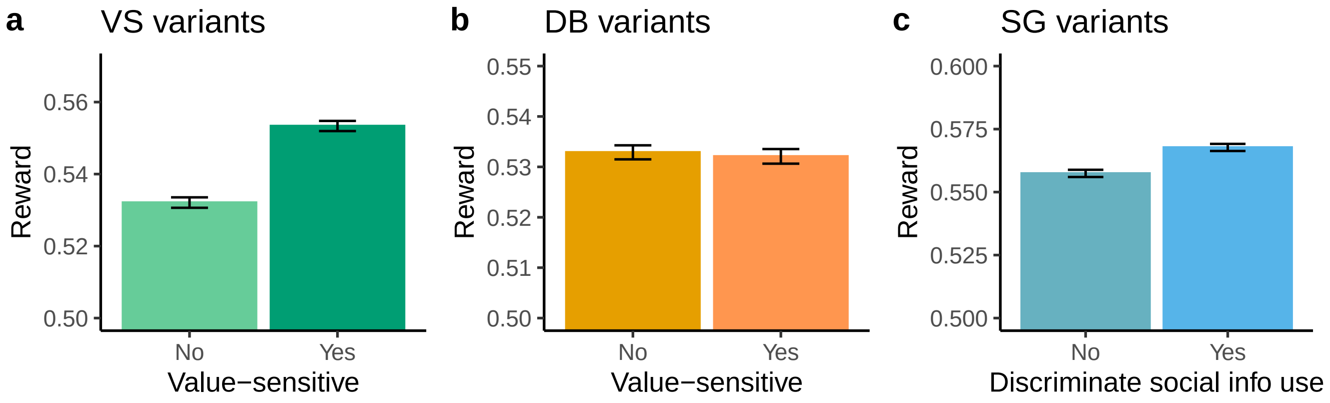

Figure S2. Model variants. a) Value Shaping with (canonical) and without value sensitivity added. **b**) Decision Biasing with and without (canonical) value sensitivity added. **c**) SG with ε_{soc} set to 0 and not (canonical).

Detailed evolutionary simulations

Visualization as a ternary plot as in the main text only allows for comparisons between 3 models at a time. As this paper put a focus on social learning, we chose to compare the three candidate social models in this manner. Figure S3a shows the evolutionary trajectories for all starting populations, including AS. As reported in the main text, SG takes over and dominates all populations.

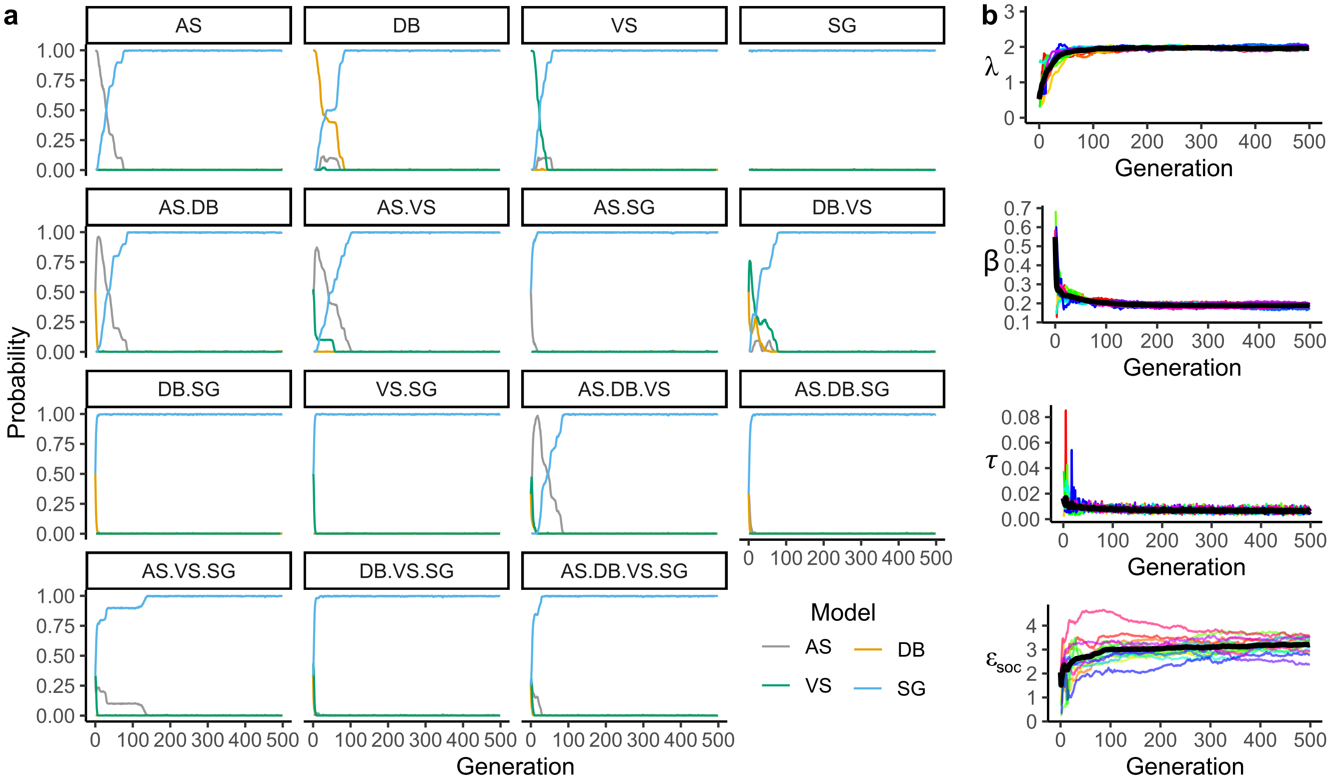

Figure S3. Evolutionary simulations in correlated environments in detail. a) Evolutionary trajectories across all starting populations. Facet labels show initial population, and lines show the probability of a given model in the population. Social Generalization dominates across all initial populations. **b**) Evolved parameters for Social Generalization. Thick black line is the average.

When it comes to parameter evolution, we can gain insight into the "optimal" SG agent based on evolutionary simulations as well (Fig. S3b). We only report SG parameters as parameters of other models are unstable due to their low population size. The parameter evolutions are discussed in the main text.

We additionally ran evolutionary simulations in a setting analogous to previous literature[2] with one expert choosing the correct option and the (social) learning agent being in an identical environment. This served to show that the spatially correlated bandit setup is not inherently different from simpler bandits used in previous literature. As reported in the main text, we replicate VS being the dominant model in such settings when comparing only the previously established models (Fig. S4a), and find an equilibrium between VS and SG when considering all of our candidate models (Fig. S4b-c). This is because when fully socially reliant (α=1 for VS, or $\varepsilon_{soc} = 0$ for SG), as is optimal when learning from an expert in the same environment, and in the same environment as said expert, VS and SG make identical choice predictions, only differing at the stage at which social information is integrated (Fig. S4d). SG can be viewed as an extension of VS to cases where one has to learn from others in non-identical environments, not a completely new model.

Reward improvement

In the analysis of experiment 1, we investigated how participants used both individual and social information to guide their exploration. To this end, we analyzed the influence of improvement potential (the difference between previous individual and previous social reward in the case of social information, and the difference between maximum possible reward and previous individual reward for individual information) on reward improvement (the difference between current and previous individual reward). While the data corroborated no effect of negative social information (improvement potential < 0), there seemed to be a strong relationship between positive social improvement potential and reward improvement (Fig. S5a). When modelling this relationship, we not only found a general relationship between improvement potential and improvement (0.53 [0.47, 0.60]), but also both a significant positive effect of social information (0.06 [0.04, 0.07]) and its interaction with improvement potential (0.12 [0.09, 0.15]; Fig. S5b). This seemed to indicate that social information was even more effective than individual

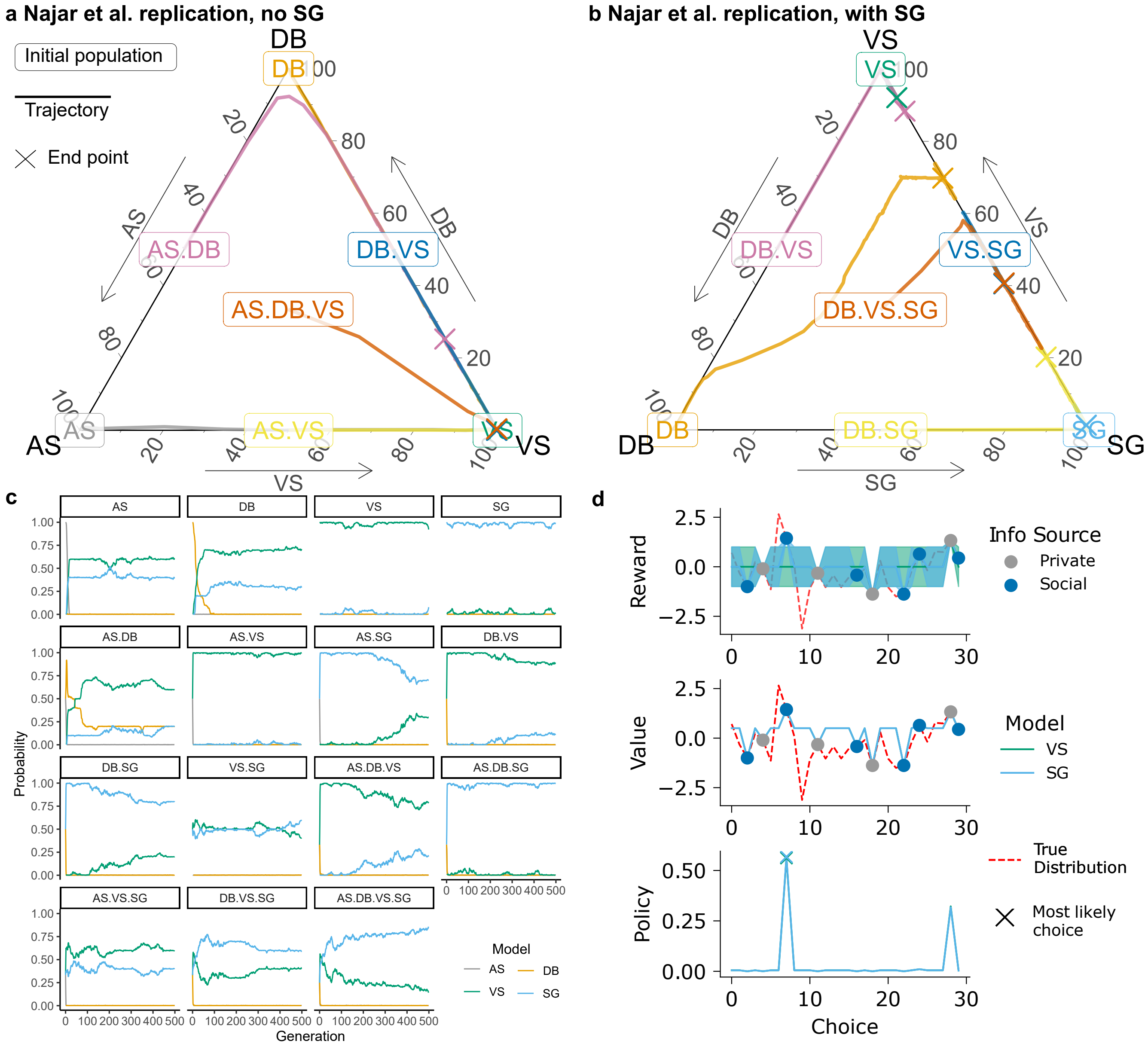

Figure S4. **Evolutionary simulations in identical environments with an expert (Najar et al. setting)[2].** **a**) Evolutionary trajectories when only considering AS, DB, and VS. **b**) Evolutionary trajectories when only considering the social models (no AS). **c**) Full evolutionary trajectories (all possible initial populations). Facets labels give initial populations, and lines show the probability of a given model. **d**) 1-dimensional illustrative example to compare VS and SG. When reward landscapes are identical, VS and SG agents which fully rely on social information ($\alpha = 1$ or $\varepsilon_{soc} = 0$) behave identically.

information in guiding participants' exploration.

However, while the relationship remained similar in experiment 2 (Fig. S5c), and the baseline effects replicated (improvement potential: 0.53 [0.51, 0.55]; social information: 0.05 [0.04, 0.06]; their interaction: 0.12 [0.09, 0.15]), we found none of their interactions with task type were significant (improvement potential*group round: 0.00 [-0.03, 0.03]; social info*group round: 0.01 [-0.01, 0.02]; improvement potential*social info*group round: -0.01 [-0.05, 0.03]; Fig. S5d). This shows that the effect we found in experiment 1 was solely based on the task structure, and not any actual benefit of social information usage.

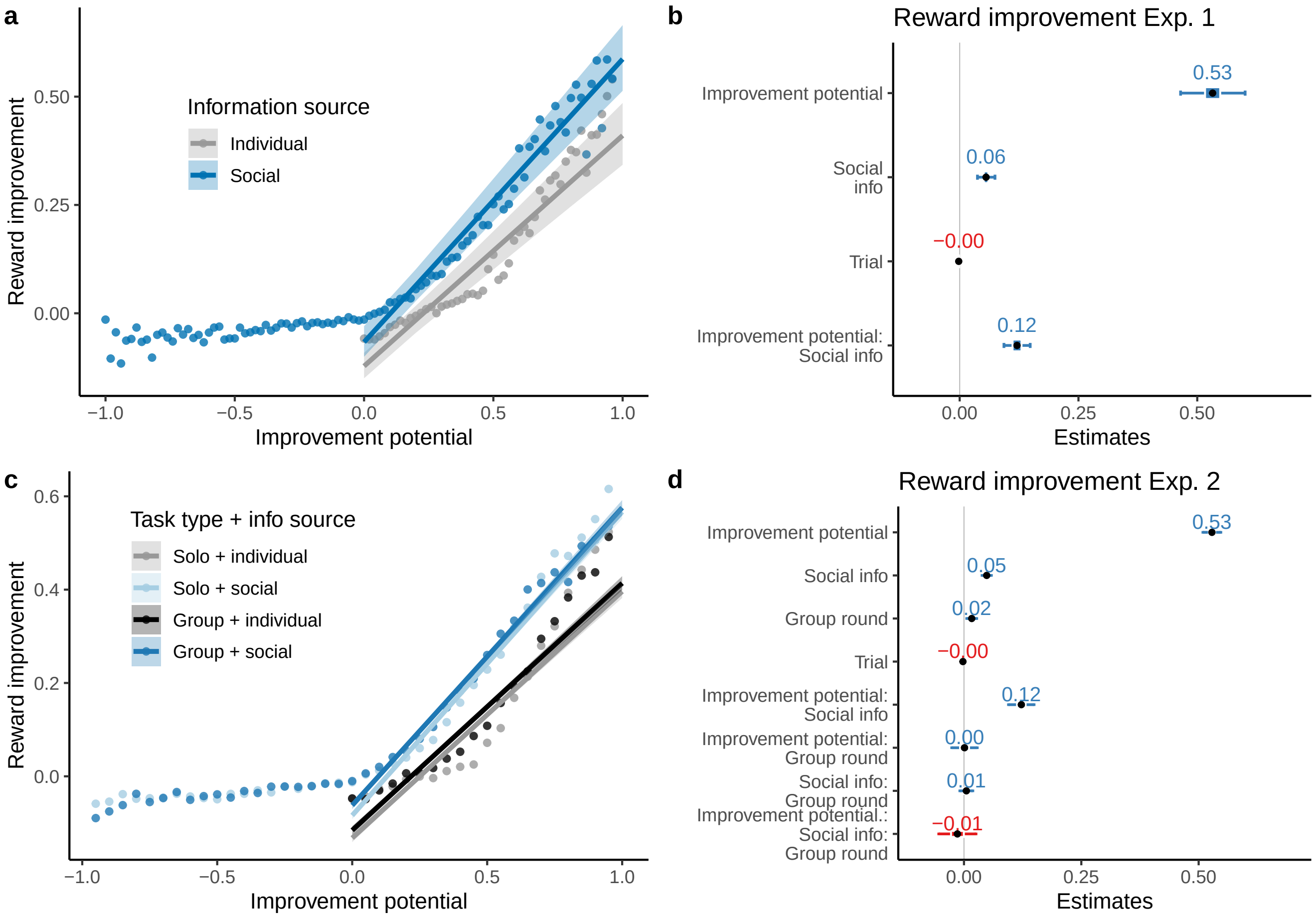

Figure S5. Reward improvement analysis for Exp. 1 (a-b) and Exp. 2 (c-d). a) Hierarchical Bayesian regression of reward improvement (current - previous individual reward) and improvement potential (social: previous social reward - previous individual reward; individual: 1 - previous individual reward) in Exp. 1. **b**) Regression parameters for the regression shown in a. **c**) Hierarchical Bayesian regression of reward improvement (current - previous individual reward) and improvement potential (social: previous social reward - previous individual reward; individual: 1 - previous individual reward) in Exp. 2. **d**) Regression parameters for the regression shown in c.

Model bounding

As the baseline asocial learning model is nested in all social models, we determined bounds for the social models to minimize model mimicry, and thus improve recovery. This serves to make the modelling more stringent compared to previous work[62]. The bounds were determined based on the social mechanisms of the respective models. Since DB effectively only changes imitation rate (mixing parameter γ effectively trades off between individual learning and imitation, which makes it interpretable as an average imitation rate per trial), we chose to determine the bound based on expected average imitation based on individual learning in correlated environments. We simulated AS with priors from previous literature, and set the lower bound at 95%-quantile of the resulting Poisson distribution based on imitation counts (Fig. S6a). This meant, that agents fit by DB were expected to imitate at least as much as the 5% tail-end of the asocial population.

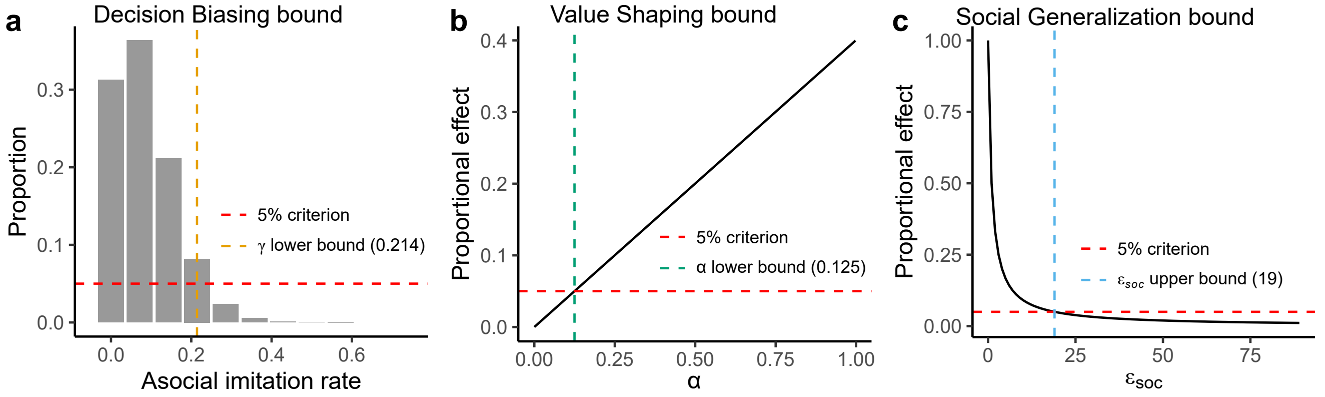

Figure S6. Model bounding of the social models. Illustrations of the model bounding concepts, criteria, and parameter cut-offs for Decision Biasing (**a**), Value Shaping (**b**), and Social Generalization (**c**)

Since VS affects the value function at a social observation, we determined the lower bound based on the minimum effect of a maximum reward social observation on a naive social learner (no individual observations) across a range of β-values. The criterion was a minimum of 5% change from the individual value (Fig. S6b).

For SG, ε_{soc} affects how strong of an effect social information has on the posterior of the GP. Hence, we set the bound at the social observation retaining at least 5% of its value given a naive social learner (Fig. S6c).

Model and parameter recovery

To assess model recovery, we simulated data using parameters fitted to participants for all models in experiment 1 and the group rounds of experiment 2. We then fit the simulated participants following the same procedure as used for actual participants, assigned each simulated participant a model based on best fit. We computed conditional probabilities for confusion and inversion matrices (Fig. S7). Despite the bounding, there is still some confusion potential between models, especially DB and VS, which hardly get fit with social parameters above the lower bound, and AS. There is also some confusion potential between AS and SG, but it is roughly balanced between the two, so overall fitting results should not be biased either way. In turn, the likelihood of a fit model being the generating one is highest for DB (0.85) and VS (0.9), but still high for AS (0.7) and SG (0.79).

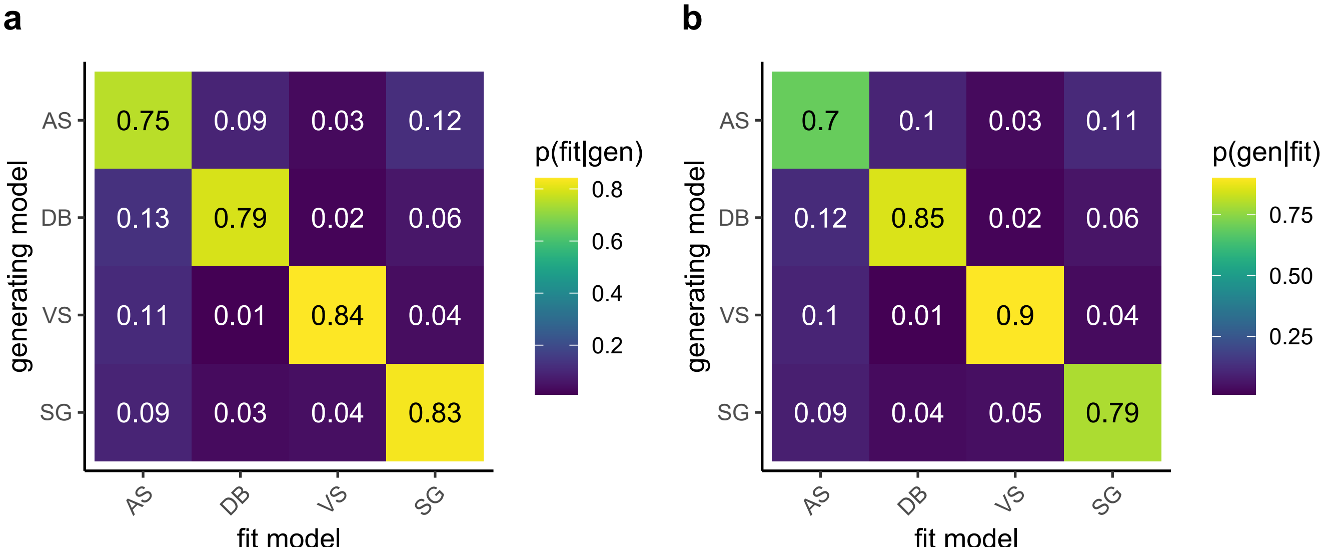

Figure S7. Model recovery. a) Confusion matrix giving the conditional probability of a model being the best fit given the generating model. Values on the diagonal give the probability that the correct model is fit. **b**) Inversion matrix giving the conditional probability of a model being the generating model given it is fit. Values on the diagonal give the probability that the fit model is correct.

When it comes to parameter recovery, we used the same procedure of simulating and fitting the data as for model recovery, but looked at the correlations between the parameters of the generating model and its fit instead of the best fitting model (Fig. S8). λ ($r_\tau = .87$, $p < .001$, $BF > 100$) and τ ($r_\tau = .86$, $p < .001$, $BF > 100$) correlate near perfectly with the generating parameters across models. When it comes to β and the social parameters, the issue of lower bound social parameters recurred, leading to worse fits for DB and VS. β correlations are still high overall ($r_\tau = .85$, $p < .001$, $BF > 100$), whereas the social parameter correlation is lower ($r_\tau = .27$, $p < .001$, $BF > 100$). However, given that neither DB nor VS fit participant data well, leading to them mimicking AS as much as possible, this lack of correlation is less concerning. Looking at only the correlation for ε_{soc}, it is noticeably higher ($r_\tau = .51$, $p < .001$, $BF > 100$).

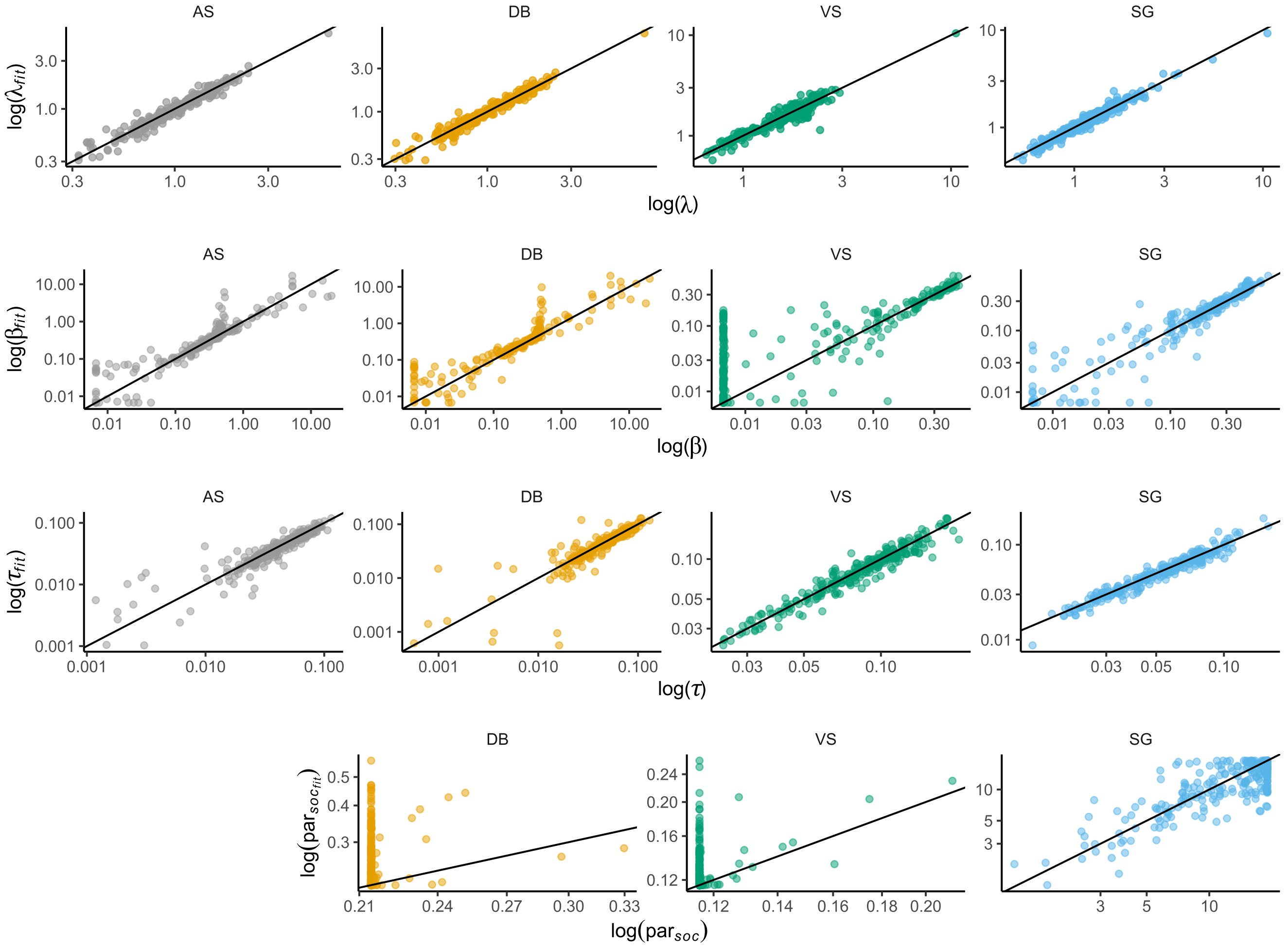

Figure S8. Parameter recovery. Relationship between generating and recovered parameter by model (facet label).

Model performance

Models were fit using leave-one-round-out cross-validation. Negative log likelihoods were summed across test-rounds, with mean values being reported in Tables S1 and S2. Pseudo-R² was computed as $R^2 = 1 - (nLL_{model}/nLL_{random})$ where nLL_{random} is treating every choice as equally likely (1/121) for all non-random trials for that participant. Random trials were excluded from model fitting.

As the models were nested, and AS generally provides a good explanation even in social settings, especially once a high value option has been found, performance does not differ greatly between models. However, SG is consistently the best fit across both experiments.

Model	Mean nLL	Pseudo-R²
AS	448.1	0.1576
DB	357.2	0.3268
VS	379.5	0.2845
SG	**356.9**	**0.3273**

Table S1. Model Performance Metrics for Exp. 1.

Model	Mean nLL	Pseudo-R²
AS	369.4	0.3106
DB	367.2	0.3147
VS	378.8	0.2932
SG	**365.5**	**0.3178**

Table S2. Model Performance Metrics for Exp. 2 group rounds.

Exp. 2 parameters

To focus on the differences between β-values, we do not report all parameter values for the group rounds of Exp. 2 in the main text (Fig. S9). Generalization parameter $\lambda \approx 1.1$, which is significantly lower than the ground truth $\lambda = 2$ ($t(52) = -14.4$, $p < .001$, $d = 2.0$, $BF > 100$). Directed exploration parameter $\beta \approx 0.22$, and random exploration parameter $\tau \approx 0.06$. Social noise $\varepsilon_{soc} \approx 9.5$, which is significantly higher than optimal the optimal value found in evolutionary simulations ($t(52) = 8.3$, $p < .001$, $d = 1.1$, $BF > 100$).

Exploration optimality

Following up the comparatively lower β-parameter in experiment 1, we compare participant's β-parameters between the solo and group rounds. As reported in the main text, participants had significantly higher β-values in solo than in group rounds ($Z = -2.7$, $p = .004$, $r = -.23$, $BF = 63$, Fig. S10a). In simulations based on participant parameters while varying the parameter of interest, we find that such low values of β are actually optimal in the current task, both in solo and group rounds. We see that the group round value of β is closer to optimal than the solo round one, and both are lower and thus closer to optimal than the average found in previous literature[40].

As we also found higher values of λ in Exp. 1, we exploratively repeat these analyses. λ is significantly higher, and thus closer to ground truth, in group than in solo rounds. Again, group round λ is closest to the theoretical optimum, followed by solo round and previous study values. Taken together, this implies that social information may improve exploration behaviour closer to optimality in general.

Exclusion analyses

To ensure that correlations were not spurious because of participants at the upper bound of ε_{soc}, we redid the correlation analyses for reward and β excluding any participants whose $\varepsilon_{soc} > 18.9999$.

All relationships remained significant. In Exp. 1, there was a significant negative correlation between ε_{soc} and mean reward ($r_\tau = -.31$, $p = .001$, $BF = 24$), indicating higher reliance on social information leading to higher scores. There was also a significant positive correlation between ε_{soc} and β ($r_\tau = .34$, $p < .001$, $BF = 61$), indicating that participants using relatively more social learning used less directed exploration and vice versa. In Exp. 2, we replicate both the relationship of ε_{soc} with reward ($r_\tau = -.20$, $p = .034$, $BF = 1.6$), and β ($r_\tau = .35$, $p < .001$, $BF > 100$).

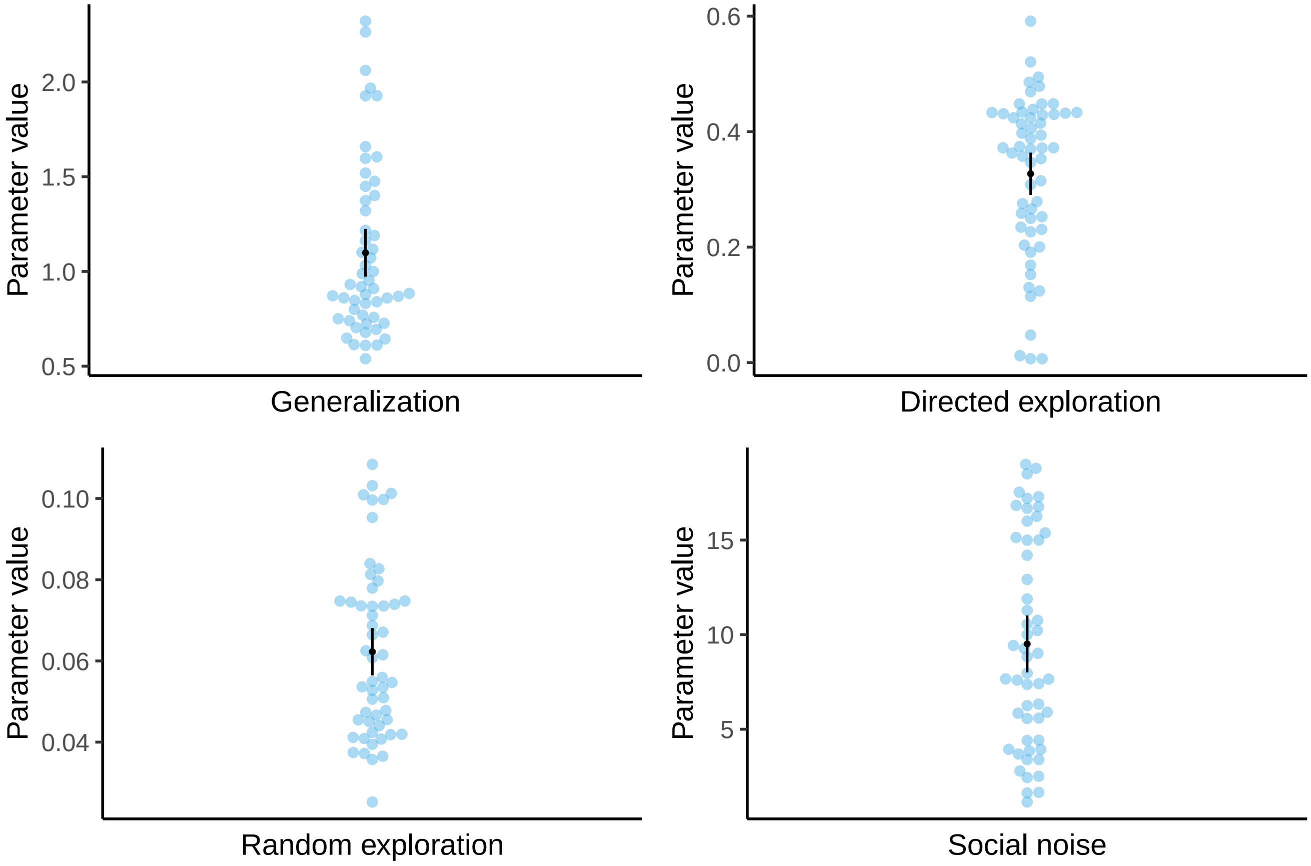

Figure S9. Parameter fits for group rounds of Exp. 2. Only participants best fit by SG shown.

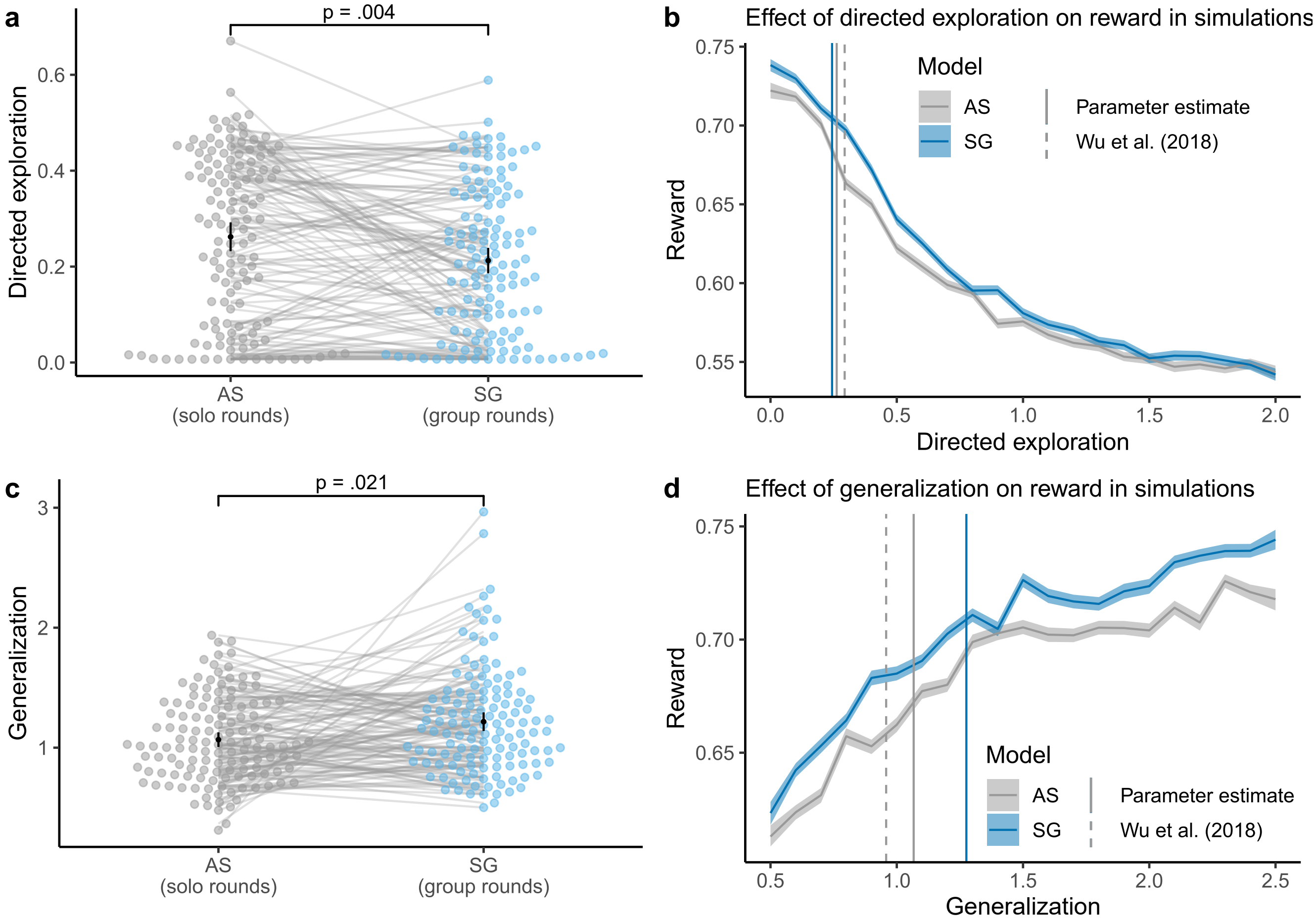

Figure S10. More optimal parameters in group rounds. a) Difference in directed exploration parameter β between participants fit by AS in solo rounds and participants fit by SG in group rounds, reproduced from main text. **b**) Relationship between directed exploration parameter β and reward in simulations. We ran simulations on the parameter ranges present in participants while varying β, to see if lower β values are more beneficial for social learners. In both models, lower values of β are associated with higher reward. Average β in group rounds is closest to optimal, followed by solo round average and average from previous literature. **c**) Difference in generalization parameter λ between participants fit by AS in solo rounds and participants fit by SG in group rounds. **d**) Relationship between generalization parameter λ and reward in simulations. We ran simulations on the parameter ranges present in participants while varying λ, to see if higher λ values are more beneficial for social learners. In both models, higher values of λ are associated with higher reward. Average λ in group rounds is closest to optimal, followed by solo round average and average from previous literature.

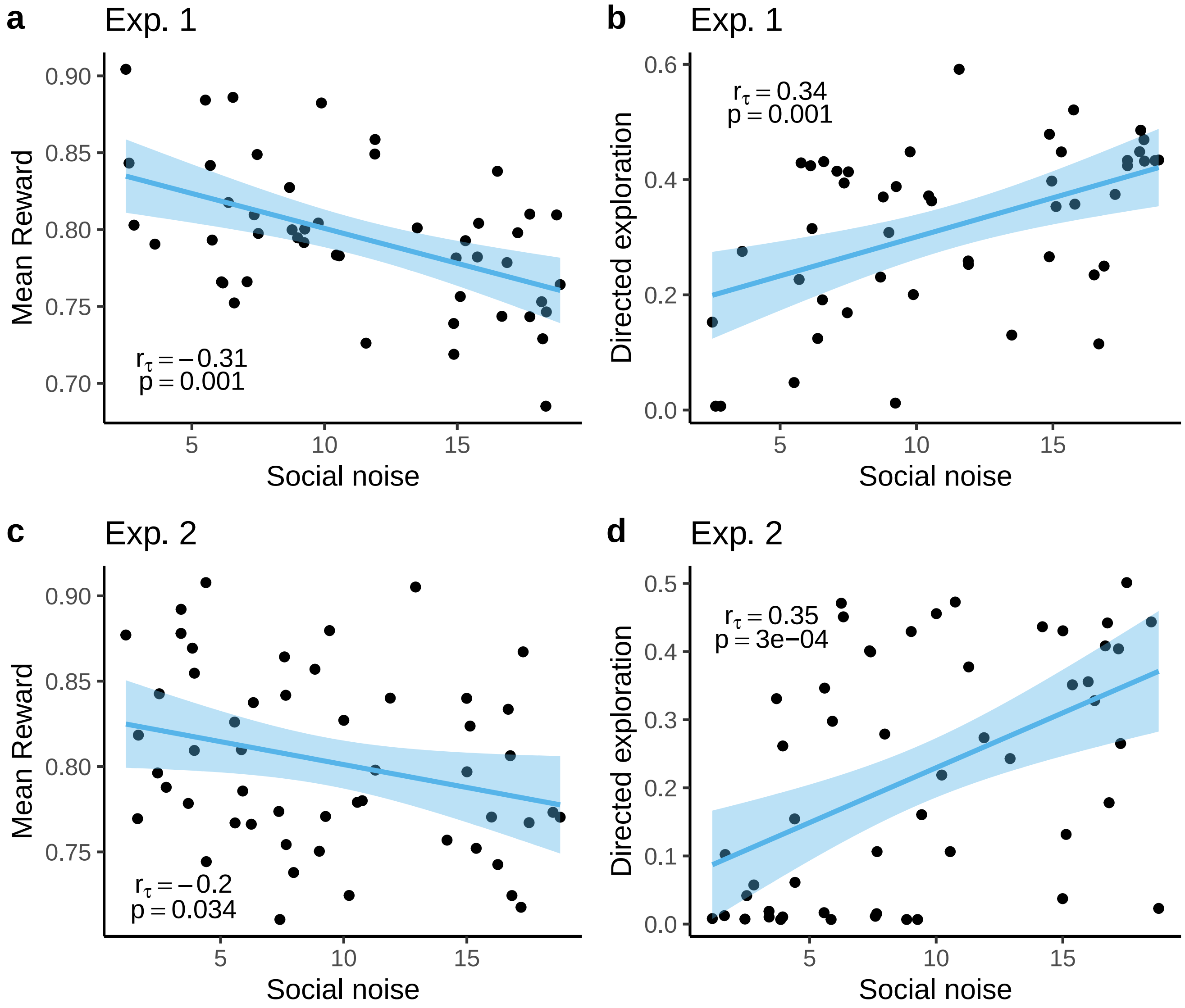

Figure S11. Correlation analyses excluding ε_{soc} at the upper bounds. a) Relationship between ε_{soc} and mean reward in Exp. 1. **b**) Relationship between ε_{soc} and β in Exp. 1. **c**) Relationship between ε_{soc} and mean reward in Exp. 2. **d**) Relationship between ε_{soc} and β in Exp. 2.