Trust Issues: Social Learning Under Misaligned Goals

Liang Lee[1,2,*] **(liang.lee@tu-darmstadt.de), Valerii Chirkov**[3,4,*] **(valerii.chirkov@hu-berlin.de), Shen Tian**[5], **Charley M. Wu**[1,2,†] **(charley.wu@tu-darmstadt.de)**, & **Alexandra Witt**[6,†] **(alexandra.witt@riken.jp)**

[1] Technical University Darmstadt, Darmstadt, Germany, [2] Hessian AI, Darmstadt, Germany, [3] Humboldt University of Berlin, Berlin, Germany, [4] Science of Intelligence Excellence Cluster, Technical University of Berlin, Berlin, Germany, [5] Karolinska Institutet, Stockholm, Sweden, [6] RIKEN Center for Brain Science, RIKEN, Wako, Japan

[*,†] Equal contribution

Abstract

Computational models of social learning often assume learners and demonstrators share identical or at least positively correlated goals. Yet this assumption limits applications to real-world scenarios, where preferences may be misaligned or even opposed. We address this gap by extending the socially correlated bandit task to settings where agents need to learn when social information is positively correlated, uncorrelated, or negatively correlated, analogous to learning whom to trust or distrust. We introduce Social Correlation–Adjusted LEarning (SCALE), a multi-output Gaussian Process model that learns the covariance structure between agents' preferences. Using simulations, we characterize the model's performance across social environments and outline a path toward agents that can dynamically infer social correlations from experience. Our model allows us to reframe prior experimental observations, and lays the groundwork for future experimental work on the integration of preferences into individual decision-making.

Keywords: social learning; generalization; agent-based modeling; trust learning

Introduction

Imagine you are going to the cinema to watch a movie. While you could choose blindly, you might also leverage social information to make a more informed choice. You can check online reviews under the assumption that they reflect the quality of the movie. But maybe you have a friend with immaculate taste from whom you would rather solicit a recommendation, expecting it to be more aligned with your taste than any random reviewer. Or perhaps you unironically enjoy "bad" movies and would rather choose the one with the lowest rating instead. In each case, the same social signal carries different informational value, depending on how others' preferences relate to one's own. Effective social learning therefore requires learning when to trust, ignore, or invert social information (Fig. 1a).

While this type of choice problem is well-studied in recommender systems, where collaborative filtering leverages large-scale datasets of user preferences to generate recommendations (Papadakis et al., 2022), such approaches typically require extensive prior data. In contrast, research on social learning focuses on sparse data settings, where agents integrate social information with their individual decision-making process, often from only one or a few demonstrators (e.g., McElreath et al., 2008; Molleman et al., 2020; Najar et al., 2020; Toyokawa et al., 2019; Wu, Deffner, et al., 2025).

However, in this endeavor to understand social learning from sparse data, much of the literature has relied on simplified settings. Most commonly, demonstrator and observer are assumed to share the same goal and value function. In such experimental settings, we commonly find under-use of social information compared to normative performance (Acerbi et al., 2016; Mesoudi, 2011; Molleman et al., 2020; Morin et al., 2021; Tump et al., 2018), while the opposite appears to be true for real-world social learning, where we find (often anecdotal) evidence for the overuse of social information (Henrich, 2015; Mesoudi, 2009). One possible explanation for this discrepancy is that humans may be adapted to learning from others with distinct preferences, which is rarely captured in experimental paradigms (although see Collette et al., 2017). When different preferences are considered, prior work has generally focused on how observers infer others' preferences, rather than on how such inferences shape the use of social information (Fujisaki et al., 2022; Tarantola et al., 2017).

Witt et al. (2024) took an important first step towards introducing differing preferences into decision-making by investigating social learning in positively correlated environments. Their Social Generalization (SG) model treats social information as uniformly noisier than individual information, yielding a normative solution under fixed positive correlations, and providing the best descriptive account of participant behavior. While these results show that people can flexibly use social information, there are two key limitations. First, correlations were fixed and positive across all participants, despite the fact that preference similarity varies widely and can even be negative, which substantially alters the optimal social learning strategy (Analytis et al., 2018). Second, SG does not explicitly model correlation structure, instead absorbing preference differences into undifferentiated noise. As a result, it remains unclear how knowledge about others' preferences should guide individual decision-making in settings where preferences may be unrelated or even misaligned.

To address this gap, we introduce *Social Correlation–Adjusted LEarning* (SCALE) as a model that explicitly represents and exploits varied correlations across demonstrators. We test it in an extension of the socially correlated bandit task (Witt et al., 2024) with differing correlations between the agent and demonstrators, describing its characteristics and performance in comparison to SG. Our work aims to elucidate the benefits of social learning in more complex, naturalistic settings than previously investigated in the literature.

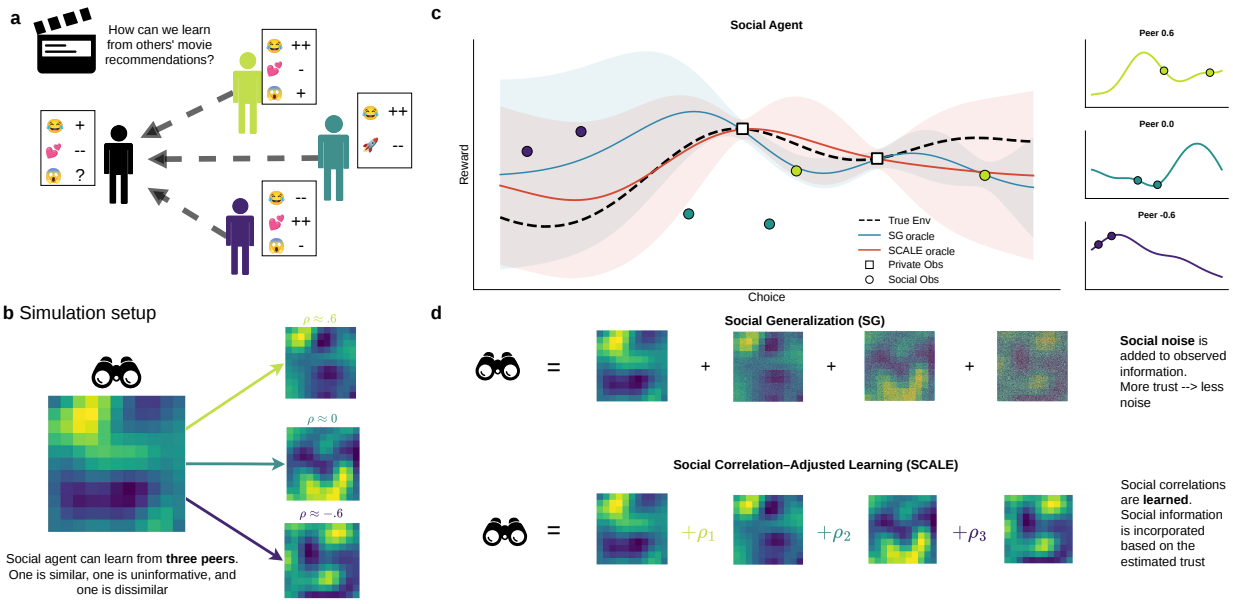

Figure 1: **a)** In real-world social learning, we often have to integrate information from others with varied preferences. **b)** Environmental structure of the task. One social learning agent observes three asocial agents with varying correlations. **c)** Differences in social information use. Given the same observational data, SG overfits to trusted agents' observations, while disregarding others. SCALE can incorporate observations from anticorrelated agents, and does not overrely on correlated observations. **d)** Concepts underlying the different models. In SG, individual and social information are treated as information sources with different noise levels. In SCALE, each agent's environment is estimated separately based on observations across all agents and a coregionalization matrix.

Methods

We conducted agent-based simulations to evaluate SCALE's performance when learning from demonstrators with different social correlations and to test whether agents can learn and transfer correlation knowledge across environments. Our simulations build on the spatially correlated bandit task (Wu et al., 2018), where agents explore a large grid with nearby options tending to have similar rewards. Because the search horizon is too short to sample all options, agents can use these spatial correlations to generalize from observed to unseen options. We then place this task in a social setting, following the socially correlated bandit (Witt et al., 2024), where multiple agents explore grids that are positively correlated with each other. In this setting, spatial correlations determine the environmental structure (for individual generalization), while social correlations determine how informative other agents' choices and outcomes are for individual decision-making (social generalization). Here, we further extend the task to heterogeneous social correlations, which can be positive, uncorrelated, or even negatively correlated (Fig. 1b).

Environment Generation

To generate socially correlated reward environments, we used a two-stage Gaussian Process (GP; Rasmussen & Williams, 2006) sampling procedure. Following Witt et al. (2024), we first sampled a parent reward function from a GP prior $f_{\text{parent}} \sim \text{GP}(0, k_{\text{RBF}}(\mathbf{x}, \mathbf{x}'))$ parameterized by a radial basis function (RBF) kernel:

$$k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\lambda^2}\right), \tag{1}$$

with length scale $\lambda = 2$ and observation noise $\sigma_\varepsilon^2 = 0.0001$. We then generated child reward functions for each agent using a Cholesky decomposition method to enforce the correlation structure. Maps were iteratively generated to ensure correlations matched the target values $\rho \in \{+0.6, 0.0, -0.6\}$ within a tolerance of ± 0.1.[1] The positive case matches previous work (Witt et al., 2024), but we add an irrelevant agent and one with symmetric inverse correlations. All reward functions were min-max normalized to the range $[0, 1]$ and defined on an 11×11 spatial grid.

Computational Models

Building on prior work (Witt et al., 2024; Wu, Meder, & Schulz, 2025), we introduce a family of GP–based reinforcement learning models that formalize how participants integrate individual and social observations, while also aligning with the environmental generation process. We present two baseline models from prior literature: Asocial Learning (AS),

[1] This iterative generation process was required because any valid correlation matrix—both the one used for environment generation and the co-regionalization matrix B in the SCALE model—must be positive semi-definite. This constraint also enforces strict bounds on the correlation ranges between agents, making the generation of environments with more extreme correlation values infeasible.

which ignores social information, and Social Generalization (SG) (Witt et al., 2024), which treats social observations as noisier but structurally identical to individual samples. Finally, we present our novel *Social Correlation–Adjusted LEarning* (SCALE) model, which exploits the environmental structure by modeling how similar each demonstrator's reward function is to the learner's own, and allows agents to learn these correlations from experience.

Asocial Learning (AS). As a baseline for performance without social influence, we modeled AS agents using GP regression. This model assumes that each agent learns a latent reward function $f(\mathbf{x})$ mapping spatial inputs $\mathbf{x} \in \mathcal{X}$ to reward observations $y \in \mathbf{y}_t$, based on its own history of observations $\mathcal{D}_t = \{X_t, \mathbf{y}_t\}$. The agent's belief about the reward landscape is represented by a posterior distribution characterized by a mean $m(\mathbf{x}^*)$ and variance $v(\mathbf{x}^*)$ for any given option $\mathbf{x}^*$:

$$\begin{aligned} m(\mathbf{x}^*) &= k_{*,t}^\top \left(K + \sigma_\varepsilon^2 I\right)^{-1} \mathbf{y}_t, \\ v(\mathbf{x}^*) &= k(\mathbf{x}^*, \mathbf{x}^*) - k_{*,t}^\top \left(K + \sigma_\varepsilon^2 I\right)^{-1} k_{*,t}. \end{aligned} \tag{2}$$

Here, $K + \sigma_\varepsilon^2 I$ represents the covariance matrix (Eq. 1) of observed locations, accounting for observation noise σ_ε^2. The term $k_{*,t}$ is the covariance vector representing the similarity between the target location $\mathbf{x}^*$ and all previously sampled locations X_t, while $k(\mathbf{x}^*, \mathbf{x}^*)$ is the prior variance at the target location. Learning uses the same RBF kernel as environment generation (Eq. 1), where larger values of λ generalize observations over greater distances.

We model the exploration–exploitation trade-off using an Upper Confidence Bound (UCB) value function:

$$V(\mathbf{x}) = m(\mathbf{x}) + \beta \sqrt{v(\mathbf{x})}, \tag{3}$$

where directed exploration parameter β determines how much weight an agent puts on exploration. Actions are selected probabilistically using a softmax rule, $\pi_{\text{ind}}(x) \propto \exp\big(V(x)/\tau\big)$, where temperature τ determines choice stochasticity.

Social Generalization (SG). The SG model (Witt et al., 2024) integrates social observations directly into the GP by assuming heteroskedasticity: socially acquired observations are assumed to be structurally identical, but noisier than individual samples (Fig. 1c blue line, d top). This is implemented by assigning observation-specific noise variances to the covariance update:

$$\sigma_\varepsilon^2 = \varepsilon_{\text{ind}} + \delta_{\text{soc}}\, \varepsilon_{\text{soc}}, \tag{4}$$

Here, ε_{ind} represents the baseline noise for individual observations, and δ_{soc} is an indicator function taking the value of 1 for social observations and 0 otherwise. A larger ε_{soc} reduces the influence of social observations by reducing their contribution to both the posterior mean estimate and the reduction of predictive uncertainty. Apart from this weighted social information integration, the SG model retains the same spatial generalization mechanisms (RBF kernel) and decision policy (UCB-softmax) as the AS model.

Social Correlation–Adjusted LEarning (SCALE). To account for diverse social environments where preferences may be misaligned, we introduce SCALE. Unlike SG, which treats all social information as uniformly noisy, SCALE embeds the reward functions of all agents into a Multi-Output Gaussian Process with a learnable co-regionalization matrix B that explicitly represents pairwise correlations. This covariance structure corresponds to the Intrinsic Coregionalization Model (Bonilla et al., 2007; Goovaerts and Goovaerts, 1997; Rasmussen and Williams, 2006).

Multi-output kernel. The covariance function in this framework factorizes into spatial and social components. For each observation at location $\mathbf{x}$ by agent i and location $\mathbf{x}'$ by agent j, the kernel is defined as:

$$k\big((\mathbf{x}, i), (\mathbf{x}', j)\big) = k_{\text{RBF}}(\mathbf{x}, \mathbf{x}')\, B_{ij}. \tag{5}$$

Here, k_{RBF} handles spatial generalization and B is the symmetric coregionalization matrix. The diagonal entries $B_{ii} = 1$ represent individual variance, while the off-diagonal entries $B_{ij} = \rho_{ij}$ encode the correlation between the learner i and demonstrator j. Correlation ρ_{ij} acts as the dynamic trust parameter updated through experience.

SCALE differs in key respects from SG's noise-based weighting: while SG can only downweight social observations, SCALE can amplify information from positively correlated peers, invert signals from negatively correlated peers, and ignore uncorrelated peers (Fig. 1c red line, d bottom). This enables more optimal use of diverse social information, rather than treating all differences as noise.

Learning trust. The learning agent estimates pairwise correlations ρ_{ij} by assessing the similarity between its own preferences and those of each demonstrator. The intuition mirrors everyday similarity judgments: if you and a friend both enjoyed *The Matrix*, did your preferences align by chance, or did you genuinely share taste? To answer this, the agent compares its own latent reward estimates to the values implied by the demonstrator's observations, specifically at the locations X_j the demonstrator has visited.

For each demonstrator j, the agent computes two posterior estimates restricted to these visited locations. First, it predicts rewards based on its own private observations $m_i(X_j)$ (see Eq. 2). Second, it estimates the rewards based solely on the demonstrator's observations, $m_j(X_j)$, effectively filtering the social data through its own GP kernel (i.e., using their own λ). This comparison allows the agent to check for alignment without needing to simulate the demonstrator's beliefs across the entire unobserved grid or having direct access to the demonstrator's internal beliefs.

The agent then computes a variance-weighted correlation between these two local estimates ($m_i(X_j)$ and $m_j(X_j)$):

$$\hat{r}^j_{\text{weighted}} = \text{Corr}_{W_j}\left(m_i(X_j), m_j(X_j)\right) \tag{6}$$

with elements weighted by $W_j = 1/(v_i(X_j) + v_j(X_j) + \varepsilon)$, which is inversely proportional to the combined predictive

uncertainty, with ε added for numerical stability. This correlation estimate serves as the evidence for updating trust. If $\hat{r}^j_{\text{weighted}}$ is positive, the agent infers similar preferences; if negative, opposing preferences; if near zero, independent preferences. The trust parameter is then updated via a delta rule with fixed learning rate α:

$$\hat{\rho}^j_{t+1} = \hat{\rho}^j_t + \alpha(\hat{r}^j_{\text{weighted}} - \hat{\rho}^j_t). \quad (7)$$

Critically, the learned $\hat{\rho}^j_t$ directly parameterizes the coregionalization matrix B_{ij} in the multi-output kernel (Eq. 5), which governs how social observations influence the agent's spatial predictions on all subsequent trials. As the agent refines its estimates of whom to trust, these beliefs automatically propagate through the GP to adjust how strongly it weighs each demonstrator's observations when generalizing across space.

Through this process, agents adaptively discover whom to trust (positive ρ), whom to distrust (negative ρ), and whom to ignore (zero ρ), starting from $\rho = 0$ with no prior knowledge and approaching the performance of agents with perfect *a priori* knowledge of the correlation structure.

Simulation Design. All simulations involved one focal agent observing three asocial demonstrators whose reward functions were correlated with the focal agent's at $\rho = \{+0.6, 0.0, -0.6\}$. In each trial, the focal agent observes both its own outcome, and all demonstrator choices and outcomes. It then uses this information to make its next choice.

We ran two simulation regimes. To assess learning within a single environment, we ran 50,000 independent simulations of 30 trials each. To assess transfer learning, we ran 50,000 multi-round chains, each consisting of 5 sequential rounds of 15 trials, with new reward landscapes resampled in each round while the correlation structure was held constant. For learning agents, final correlation estimates from round r served as initial priors for round $r+1$. Model parameters were sampled from priors following Witt et al. (2024) when available: length scale $\lambda \sim \text{LogNormal}(-0.75, 0.5)$, exploration parameter $\beta \sim \text{LogNormal}(-0.75, 0.5)$, temperature $\tau \sim \text{LogNormal}(-4.5, 0.9)$; otherwise, for all simulations we fix observation noise $\sigma^2_\varepsilon = 0.1$ and correlation learning rate $\alpha = 0.05$ for stability. We additionally conducted a sensitivity analysis varying $\alpha \in \{0.01, 0.02, 0.05, 0.10, 0.20\}$ and $\sigma^2_\varepsilon \in \{0.010, 0.032, 0.100, 0.316, 1.000\}$ to assess robustness of the main findings.

Results

We evaluate AS, SG, and SCALE focusing on three questions: (1) How much benefit does explicitly modeling heterogeneous correlation structure provide vs. treating all social information as uniformly noisy? (2) Can agents adaptively infer social correlations from experience without prior knowledge? (3) How does performance change across multiple rounds when correlation estimates carry over?

Agents. We test seven different agent types. **AS** is a baseline agent that ignores all social information. **SG fixed** is the model as described by Witt et al. (2024), with social observation noise stable for all demonstrators ($\varepsilon_{\text{soc}} = \{3, 3, 3\}$). To compare best- and worst-case versions of SG, we define **SG oracle** as an agent with observation noise parameters tuned to appropriately weight the three demonstrators ($\varepsilon_{\text{soc}} = \{0.01, 20, 20\}$), while **SG anti-oracle** used inverted noise parameters ($\varepsilon_{\text{soc}} = \{20, 0.01, 0.01\}$). Analogously, we use **SCALE oracle** as an oracle agent with perfect knowledge of the true correlation structure ($\rho = \{+0.6, 0.0, -0.6\}$), while **SCALE anti-oracle** is an agent with inverted beliefs about correlation structure ($\rho = \{-0.6, 0.0, +0.6\}$), representing maximally miscalibrated social learning. Lastly, our **SCALE** model defines an agent that adaptively infers correla-

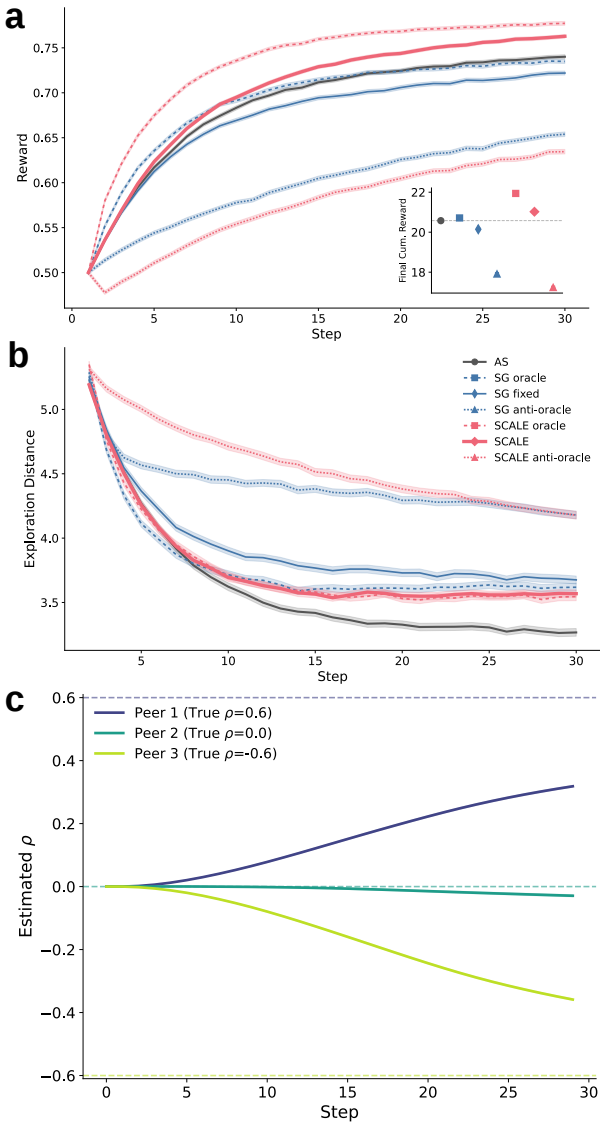

Figure 2: Performance and learning dynamics of the seven candidate models. **a)** Mean cumulative reward over time (main plot) and total performance (inset). **b)** Exploration distance over time, measured as Euclidean distance between consecutive choices. **c)** Estimated correlation parameters ($\hat{\rho}$) for the SCALE model. Dashed lines indicate true values.

tions from observed alignment between predictions and social observations (initialized at $\rho = 0$ for all demonstrators).

Calibrated Beliefs Enable Optimal Social Learning. SCALE oracle achieved the highest performance across all agents, while SG oracle achieved similar performance to AS (Fig. 2a). SCALE achieved the second-best performance, outperforming SG oracle from around trial 10 and approaching the performance of the SG oracle model near the end (Fig. 2a inset). Both anti-oracle models performed similarly poorly, confirming how miscalibrated weighting schemes, whether implemented via observation noise or correlation structure, lead to systematic exploitation of misleading information, with greater losses by flipping the correlation structure (SCALE anti-oracle).

Exploration and Exploitation Dynamics. To understand these performance results in more detail, we next investigate the exploration dynamics of the different models. All models exhibited a characteristic decline in exploration distance as agents transitioned from exploration to exploitation (Fig. 2b). However, the rate and asymptotic level of this decline varied across models, which explains their relative performance. Well-calibrated models (AS, SG and SCALE oracle, and SCALE) identified high-value regions more efficiently and subsequently exploited them, while miscalibrated models (SG and SCALE anti-oracle) maintained higher exploration distances throughout the session, reflecting persistent uncertainty due to contradictory social information that misled the agent away from promising areas of the search space.

Learning Trust from Experience. Finally, we investigate how SCALE approaches oracle-level performance over time. Although it starts with no prior knowledge of demonstrator correlations ($\rho_{\text{init}} = 0$), the agent's estimated correlation ($\hat{\rho}$) systematically approached the true values over time (Fig. 2c). The mean absolute error (MAE) of correlation estimates declined systematically from approximately 0.4 at trial 1 (when initialized at $\rho = 0$) to approximately 0.2 by trial 30. Although estimates exhibited some variability and did not fully converge to the true values within 30 trials, they were sufficient to support near-optimal decision-making, explaining the comparable performance of the calibrated variant SCALE oracle and SCALE at the end of the learning process.

Transfer of Learned Trust. While SCALE oracle naturally outperforms SCALE in one-shot scenarios, we often learn from the same group of people repeatedly, which makes it worthwhile to invest effort into understanding demonstrator preferences for higher payoffs in the future. To test this intuition, we investigate SCALE performance across different reward environments with the same correlation structure across demonstrators. We conducted a multi-round simulation with 50,000 independent chains, each consisting of 5 sequential rounds of 15 trials. In each round, a new environment was generated with different reward functions, but a constant correlation structure ($\rho = \{+0.6, 0.0, -0.6\}$). For the SCALE model, the final estimated correlation parameters from round

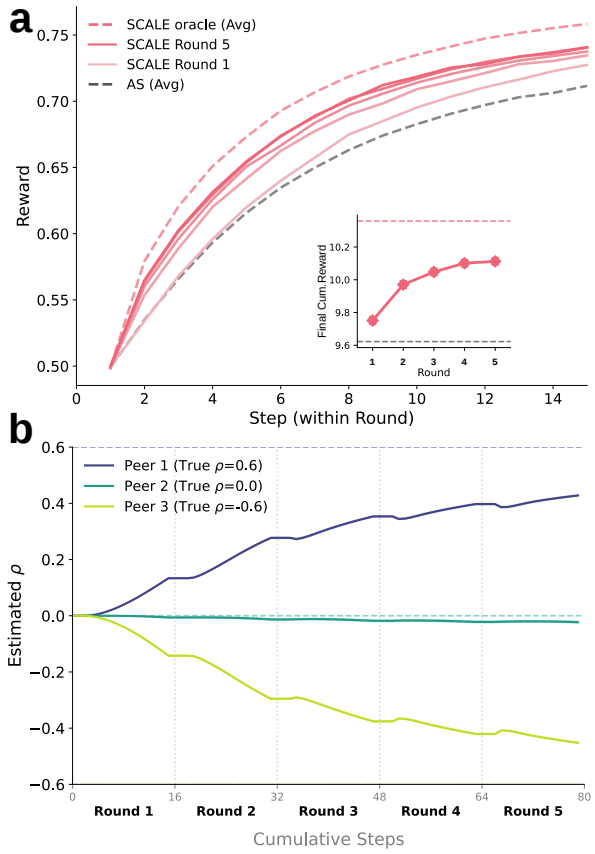

Figure 3: Transfer learning across environments. **a)** Learning dynamics within round (main plot) and total performance across rounds (inset). Error bars represent 95% confidence intervals. **b)** Evolution of correlation estimates ($\hat{\rho}$) for the three demonstrators (colored by correlation).

r were used as the initial priors for round $r+1$, simulating an agent that carries forward its beliefs about demonstrator reliability across contexts. Baseline conditions (SCALE oracle and AS) maintained fixed parameters throughout, and showed stable performance.

As expected, SCALE exhibits substantial improvement over successive rounds as seen in an increase in final cumulative rewards (Figure 3a). By round 5, the learning agent's performance starts to plateau, notably closer to that of SCALE oracle than in round 1.

MAE of correlation estimates at the start of each round declined systematically from approximately 0.4 in round 1 (when initialized at $\rho = 0$) to approximately 0.25 by round 5, indicating that agents developed increasingly accurate priors with experience. However, we observed that these estimates plateaued after round 5 without further improvement. This ceiling effect likely arises from learning fluctuations within individual rounds, where noisy trial-by-trial observations prevent the estimate from fully stabilizing. Decoupling trust updates onto a slower timescale could potentially dampen these fluctuations, enabling more precise long-term convergence (see Discussion). This pattern demonstrates that cor-

relation learning exhibits a classic transfer learning profile: knowledge acquired in one environment accelerates learning in subsequent environments that share the same underlying structure.

Thus, we show that learning the social correlation structure can be advantageously transferred across multiple rounds with the same agents. This shows a benefit to SCALE despite the increased cognitive load early on when learning from the same social partners repeatedly, and potentially across correlated domains. The ability to learn whom to trust or distrust in one domain and apply that knowledge to another represents a critical form of meta-learning that enables efficient social information use.

Discussion

We introduced Social Correlation–Adjusted LEarning (SCALE) as a computational model that explicitly represents heterogeneous correlation structure among demonstrators. Through simulations, we show how agents can learn correlations from experience, which improves their performance compared to models from prior literature, especially with repeat interactions. These findings extend prior work on social generalization (Witt et al., 2024) by showing how agents can adaptively navigate social information landscapes where different sources vary systematically in their relevance to the learner's preferences.

Our key contribution is the novel SCALE model, which allows for the integration of varied preferences to learn from with interpretable parameters. Our simulation results also help reframe the previously observed under-reliance on social information in laboratory settings (Molleman et al., 2020; Morin et al., 2021) by showing that conservative social learning may be adaptive under uncertainty.

While a model of social generalization (SG) from previous literature (Witt et al., 2024) performs well when all demonstrators have the same positive correlation, we show that this performance does not carry over to settings with more diverse social correlations. In contrast, SCALE can exploit even negative correlation structures, and provides a framework with clearly interpretable parameters for studying diverse relationships between social peers. However, we do also find similar exploration patterns in oracle models of both types, supporting the hypothesis that social information may be used as an exploration device (Wu, Deffner, et al., 2025), thus offloading otherwise costly computations (Cogliati Dezza et al., 2019; Wu et al., 2022) on others.

When learning the correlation structure, SCALE agents estimate the demonstrator's posterior using their own generalization parameter λ. This connects our work to simulation theory of mind (Shanton & Goldman, 2010), which posits that we infer the internal states of others by simulating our own internal states if we were in their position. Future work could investigate the cost-benefit trade-off of this heuristic vs. trying to infer the other agent's degree of generalization, and test whether this assumption holds in real human behavior.

We currently present only simulation-based results, making empirical validation a critical next step. Behavioral experiments should test whether humans can learn and transfer correlation structure as SCALE predicts, and whether the model captures human performance in environments with heterogeneous demonstrators. Our learning mechanism assumed a specific functional form (alignment-based updating with fixed learning rates), but alternative mechanisms warrant comparison. For instance, people might employ categorization strategies (e.g., Davis et al., 2026) or learn predictive features of trust (Schultner et al., 2025; Smith et al., 2023), rather than on an individual basis. Testing these alternatives on human data would identify which best captures the social learning of trust.

While aggregate performance remained robust across simulations, individual learning trajectories revealed a fundamental speed-accuracy trade-off: high learning rates offer rapid adaptation but render correlation estimates vulnerable to transient noise (Wickelgren, 1977). SCALE outperformed AS across all tested combinations of learning rate and observation noise, confirming that the main findings are not contingent on the specific hyperparameter values chosen. To resolve the speed-accuracy tension, future work could address the stability-plasticity dilemma through time-scale separation, which is a ubiquitous principle across neuroscience and reinforcement learning (Konda & Tsitsiklis, 1999; McClelland et al., 1995; Purcell & Kiani, 2016). By decoupling slow trust updates from fast choice dynamics, agents can prevent short-term fluctuations from destabilizing long-term beliefs about whom to trust. More broadly, related work (Ten et al., 2025) has linked σ_ϵ^2 to Tikhonov regularization (Tikhonov, 1977), offering the intuition that it prevents the model from overfitting to observed data and taking social information with "a grain of salt" (Witt et al., 2024).

Future studies should also examine how trust learning interacts with participants' prior beliefs, such as optimism or pessimism about trusting others (Schulz et al., 2025), which may be beneficial in certain environments. Here, we also assumed correlation structures remain constant across environments, which is appropriate for stable interpersonal relationships, but unrealistic if preferences shift over time or vary by context (Schakowski et al., 2026; Wu, Deffner, et al., 2025). Extending SCALE to handle time-varying correlations would increase ecological validity. Finally, our demonstrators were asocial, but real social learners interact with others who themselves learn socially, creating complex interdependencies. Examining SCALE in networks of mutually learning agents represents an important extension.

In sum, social learning effectiveness depends critically on knowing whom to learn from. We showed that agents can learn this from experience and apply it across contexts, providing a normative foundation for understanding adaptive social learning in realistic environments where preferences vary systematically across individuals.

Code and Data Availability

All materials required to replicate the results are publicly available at https://doi.org/10.5281/zenodo.20053032.

Acknowledgments

We thank the Computational Summer School on Modeling Social and Collective Behaviour (COSMOS), supported by the RIKEN CBS–Toyota Collaboration Center (RIKEN BTCC), where the initial ideas for this project were developed. LL and CMW are supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (C^4: 101164709), the Hessian research funding programme LOEWE/4b//519/05/01.002(0022)/119, the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", Project No. 533717223), and the Excellence Cluster "Reasonable AI" by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy – EXC-3057. VC was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence".

References

Acerbi, A., Tennie, C., & Mesoudi, A. (2016). Social learning solves the problem of narrow-peaked search landscapes: Experimental evidence in humans. *Royal Society open science*, *3*(9), 160215.

Analytis, P. P., Barkoczi, D., & Herzog, S. M. (2018). Social learning strategies for matters of taste. *Nature human behaviour*, *2*(6), 415–424.

Bonilla, E. V., Chai, K., & Williams, C. (2007). Multi-task Gaussian Process Prediction. *Advances in Neural Information Processing Systems*, *20*.

Cogliati Dezza, I., Cleeremans, A., & Alexander, W. (2019). Should we control? the interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of Experimental Psychology: General*, *148*(6), 977.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *Elife*, *6*, e29718.

Davis, I., Jara-Ettinger, J., & Dunham, Y. (2026). Inferring the internal structure of groups through the integration of statistical learning and causal reasoning. *Nature Communications*. https://doi.org/10.1038/s41467-026-68754-0

Fujisaki, I., Honda, H., & Ueda, K. (2022). A simple cognitive method to improve the prediction of matters of taste by exploiting the within-person wisdom-of-crowd effect. *Scientific Reports*, *12*(1), 12413.

Goovaerts, P., & Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.

Henrich, J. (2015). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter. In *The secret of our success*. princeton University press.

Konda, V., & Tsitsiklis, J. (1999). Actor-Critic Algorithms. *Advances in Neural Information Processing Systems*, *12*.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McElreath, R., Bell, A. V., Efferson, C., Lubell, M., Richerson, P. J., & Waring, T. (2008). Beyond existence and aiming outside the laboratory: Estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3515–3528.

Mesoudi, A. (2009). The cultural dynamics of copycat suicide. *PLoS One*, *4*(9), e7252.

Mesoudi, A. (2011). An experimental comparison of human social learning strategies: Payoff-biased social learning is adaptive but underused. *Evolution and Human Behavior*, *32*(5), 334–342.

Molleman, L., Tump, A. N., Gradassi, A., Herzog, S., Jayles, B., Kurvers, R. H., & van den Bos, W. (2020). Strategies for integrating disparate social information. *Proceedings of the Royal Society B*, *287*(1939), 20202413.

Morin, O., Jacquet, P. O., Vaesen, K., & Acerbi, A. (2021). Social information use and social information waste. *Philosophical Transactions of the Royal Society B*, *376*(1828), 20200052.

Najar, A., Bonnet, E., Bahrami, B., & Palminteri, S. (2020). The actions of others act as a pseudo-reward to drive imitation in the context of social reinforcement learning. *PLoS biology*, *18*(12), e3001028.

Papadakis, H., Papagrigoriou, A., Panagiotakis, C., Kosmas, E., & Fragopoulou, P. (2022). Collaborative filtering recommender systems taxonomy. *Knowledge and Information Systems*, *64*(1), 35–74.

Purcell, B. A., & Kiani, R. (2016). Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proceedings of the National Academy of Sciences*, *113*(31), E4531–E4540. https://doi.org/10.1073/pnas.1524685113

Rasmussen, C., & Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Schakowski, A., Deffner, D., Kortet, R., Niemelä, P. T., Kavelaars, M. M., Monk, C. T., Pykälä, M., & Kurvers, R. H. J. M. (2026). High-precision tracking of human foragers reveals adaptive social information use in the wild. *Science*, *391*(6784), eady1055. https://doi.org/10.1126/science.ady1055

Schultner, D., Molleman, L., & Lindström, B. (2025). Feature-based reward learning shapes human social learn-

ing strategies. *Nature Human Behaviour*, *9*(10), 2183–2198.

Schulz, L., Streicher, Y., Schulz, E., Bhui, R., & Dayan, P. (2025). Mechanisms of mistrust: A bayesian account of misinformation learning. *PLOS Computational Biology*, *21*(5), e1012814.

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 527–538.

Smith, A. L., Heuschkel, S., Keplinger, K., & Wu, C. M. (2023). Constructing and deconstructing bias: Modeling privilege and mentorship in agent-based simulations. In L. Hunt, C. Summerfield, T. Konkle, E. Fedorenko, & T. Naselaris (Eds.), *Proceedings of the 2023 Conference on Cognitive Computational Neuroscience*. https://doi.org/10.48550/arXiv.2304.02351

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications*, *8*(1), 817.

Ten, A., Sakaki, M., Breit, S., Chandrasekaran, A., Murayama, K., & Wu, C. M. (2025). In search of lost memories: Modeling exploration with forgetful generalization. *PsyArxiv*. https://doi.org/0.31234/osf.io/hupq5_v1

Tikhonov, A. N. (1977). Solutions of ill posed problems.

Toyokawa, W., Whalen, A., & Laland, K. N. (2019). Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behaviour*, *3*(2), 183–193.

Tump, A. N., Wolf, M., Krause, J., & Kurvers, R. H. (2018). Individuals fail to reap the collective benefits of diversity because of over-reliance on personal information. *Journal of the Royal Society Interface*, *15*(142), 20180155.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

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

Wu, C. M., Deffner, D., Kahl, B., Meder, B., Ho, M. K., & Kurvers, R. H. (2025). Adaptive mechanisms of social and asocial learning in immersive collective foraging. *Nature communications*, *16*(1), 3539.

Wu, C. M., Meder, B., & Schulz, E. (2025). Unifying principles of generalization: Past, present, and future. *Annual Reviews of Psychology*, *76*, 275–302. https://doi.org/10.1146/annurev-psych-021524-110810

Wu, C. M., Schulz, E., Pleskac, T. J., & Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific Reports*, *12*, 1–14. https://doi.org/10.1038/s41598-022-07901-1

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature human behaviour*, *2*(12), 915–924.