# Visual-spatial dynamics drive adaptive social learning in immersive environments

**Charley M. Wu[1,2,3+], Dominik Deffner[2,4], Benjamin Kahl[2], Björn Meder[2,5], Mark H. Ho[6*], and Ralf H.J.M. Kurvers[2,4*]**

[1]University of Tübingen, Tübingen, DE
[2]Max Planck Institute for Human Development, Berlin, DE
[3]Max Planck Institute for Biological Cybernetics, Tübingen, DE
[4]Exzellenzcluster Science of Intelligence, Technische Universität Berlin, DE
[5]Institute for Mind, Brain and Behavior, Department of Psychology, Health and Medical University, Potsdam, DE
[6]Princeton University, Princeton NJ
[*]These authors contributed equally to this work
[+]Corresponding author: charley.wu@uni-tuebingen.de

## ABSTRACT

Humans are uniquely capable social learners[1]. Our capacity to learn from others across short[2] and long timescales[3] is a driving force behind the success of our species[4,5]. Yet there are seemingly maladaptive patterns of human social learning, characterized by both overreliance[6,7] and underreliance[8] on social information. Recent advances in animal research have incorporated rich visual[9] and spatial dynamics[10] to study social learning in ecological contexts, showing how simple mechanisms can give rise to intelligent group dynamics. However, similar techniques have yet to be translated into human research, which additionally requires integrating the sophistication of human individual and social learning mechanisms. Thus, it is still largely unknown how humans dynamically adapt social learning strategies to different environments and how group dynamics emerge under realistic conditions. Here, we use a collective foraging experiment in an immersive Minecraft environment to provide unique insights into how visual-spatial interactions give rise to adaptive, specialized, and selective social learning. Our analyses show how groups adapt to the demands of the environment through specialization rather than homogeneity and through the adaptive deployment of selective imitation rather than indiscriminate copying. We test these mechanisms using computational modeling, providing a deeper understanding of the cognitive mechanisms that dynamically influence social decision-making in ecological contexts. Specialization of learning strategies and selective attention thus offer insights into the adaptive foundations of human social learning.

Humans have a unique capacity for social learning that differentiates us from other animals[1,2]. We are remarkably flexible in how we learn from others[11–13], dynamically integrate personal and social information[14,15], and selectively favor social learning when our own capabilities seem lacking[16,17]. Yet there are seemingly maladaptive patterns of human social learning, characterized by overreliance on uninformed individuals, leading to information cascades and herding behavior[6,7], but also underreliance and inefficient use of beneficial social information[8]. Recent advances in animal research have incorporated rich behavioral data from visual field analysis[9], spatial trajectories[10], and network dynamics[18] to show how simple (and sometimes seemingly maladaptive) social learning mechanisms can give rise to intelligent behavior in dynamic and ecological environments. Yet similar approaches have yet to be translated into human research, which additionally requires incorporating the complexities of human learning mechanisms (e.g., strategic exploration[19] and model-based social learning[20,21]). Thus, a better understanding of human social learning in dynamic and ecological contexts can reveal the extent of our adaptivity and the specific mechanisms that give rise to group dynamics.

Here, we use an immersive collective foraging task (Fig. 1a-d) to study how people dynamically adapt their social learning strategies to different resource distributions (random vs. smooth; Fig. 1e). The virtual environment imposed a limited field of view, creating a natural trade-off between allocating visual attention towards individual search or towards peers for social imitation. We used a novel method for automating the transcription of visual field data (Fig. 1c; see Methods) in order to identify which participants and which elements of the environment were visible at any point in time. Combining visual field analysis with high-resolution spatial trajectories, we show how people dynamically *adapt* their social learning strategies to both the environments and individual performance, *specialize* their learning strategies within groups, and *selectively* direct social learning towards successful individuals (Figs. 2). Our behavioral analyses capture both the dynamics and structure of social interactions (Fig. 3), where we directly tested different learning mechanisms using computational models predicting sequential foraging decisions (Fig. 4). All results are compared to an asocial baseline, allowing us to specify our findings as uniquely social phenomena.

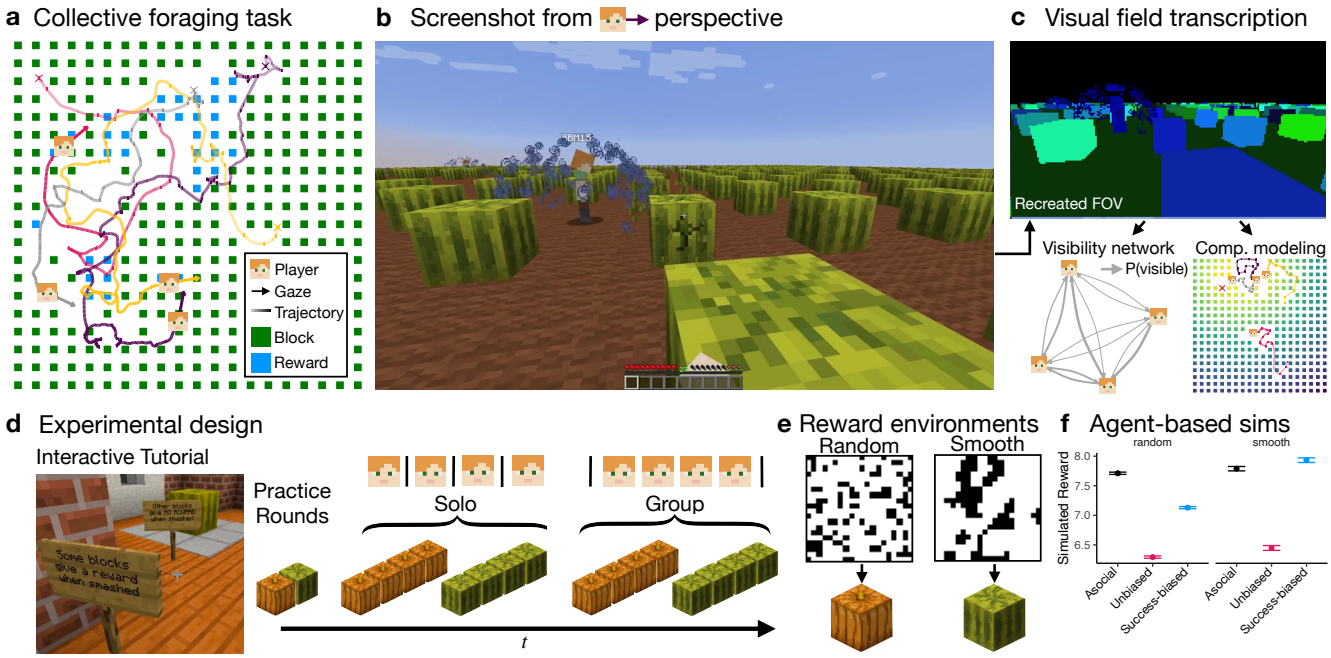


**Figure 1. Collective foraging task implemented in the Minecraft game engine**. (**a**) Participants foraged for hidden rewards in a field with 20x20 resource blocks. Each round took 120 seconds, with players starting from random locations (crosses) and gaze directions (arrows). (**b**) Screenshot from a player's perspective. Rewards (blue splash) are visible to other players, providing relevant social information for predicting nearby rewards in smooth—but not random—environments (Panel e). (**c**) Automated transcription of each player's field of view (FOV) used in visibility and model-based analyses (see Methods). (**d**) Participants learned about the task in an interactive tutorial (Supplementary Video 1) before completing two practice rounds. The main experiment consisted of 16 rounds (counterbalanced order), manipulated across condition (solo vs. group) and reward structure (random vs. smooth) with four consecutive rounds of the same type (Supplementary Videos 2-4). (**e**) Random environments had uniformly sampled rewards, smooth environments had spatially clustered rewards. Each black pixel indicates a reward from a representative sample, with both environments having the same base rate $p(\text{reward}) = .25$. The mapping to pumpkins and watermelons were counterbalanced between sessions. (**f**) Agent-based simulations (see Methods) show a benefit for success-biased social learning over asocial learning in smooth, but not random environments, whereas unbiased social learning performs poorly in both.

## Collective foraging in a virtual environment

Participants ($n = 128$) foraged for hidden rewards either alone or in groups of four, where we manipulated the resource distribution (random vs. smooth) to modify the value of social learning (Fig. 1e). Smooth environments had clustered rewards, making social observations of successful individuals (visible as a blue splash; Fig. 1b) predictive of other rewards nearby. In contrast, unpredictable rewards in random environments offered no benefits for social learning. Agent-based simulations (see Methods) demonstrate this intuition, with unbiased social learners performing poorly due to maladaptive herding, whereas selective, success-biased social learners performed better than purely asocial learners in smooth but not random environments (Fig. 1f). Thus, peers can be valuable sources of social information (in smooth environments) but also competitors for the same limited resources[11,22] (Fig. S1), creating similar real-world dynamics as developing marketplace innovations or engaging in scientific research[23].

## Temporal dynamics

To analyze the dynamics of visual-spatial interactions, we leveraged the high-temporal resolution of reward, spatial proximity, and visibility data (both seeing others and being seen) to search for temporally predictive clusters (Fig. 2a-c; see Methods). More precisely, we computed correlations between time-series at different temporal offsets (with multiple forms of chance correction), where significant clusters (bold lines) at negative offsets indicate that reward predicts future proximity/visibility and clusters at positive offsets indicate that proximity/visibility predicts future rewards.

First, the dynamics of reward (i.e., foraging success) and spatial proximity (to other players) revealed a unique pattern of *performance-adaptive spatial cycling* in smooth (but not random) environments (Fig. 2a). The negative correlation at offset -20s to -9s (bold line) indicates that poor rewards predicted increased future proximity, whereas high rewards predicted increased social distance. Additionally, the positive correlation at offset 2s to 19s indicates that high spatial

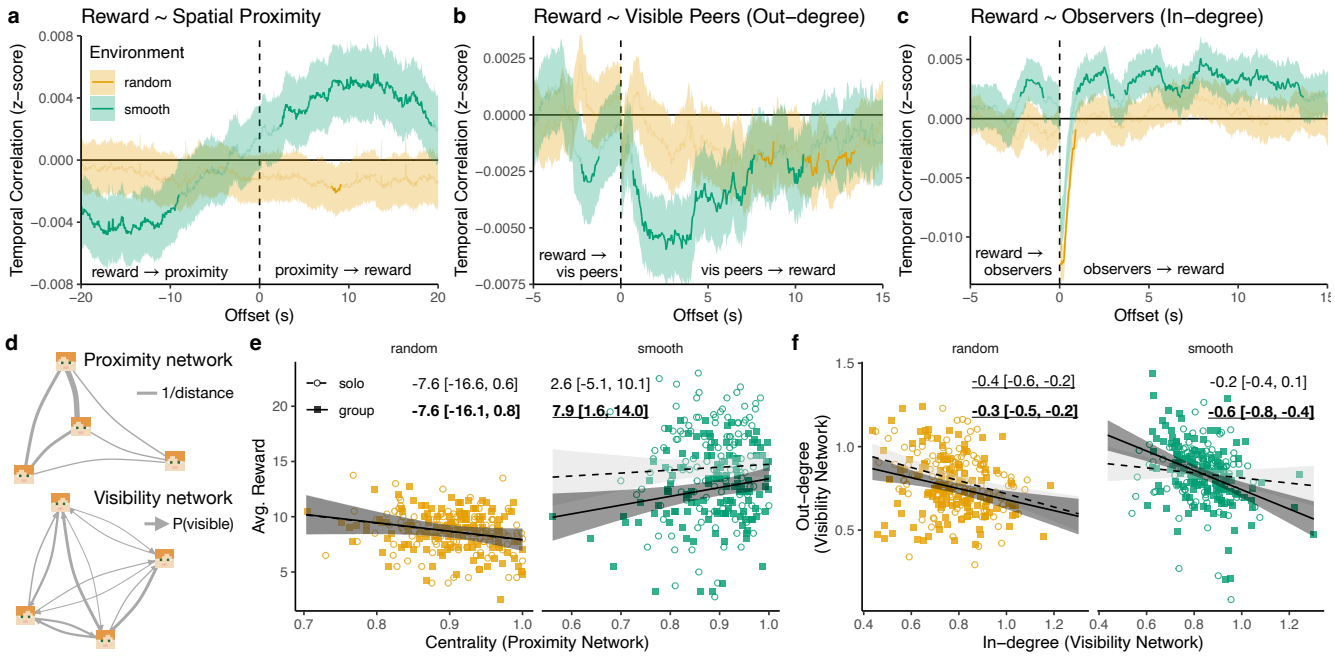


**Figure 2. Behavioral results**. (**a-c**) Temporal dynamics of reward rate and proximity/visibility in group rounds (see Fig. S3 for comparison to solo rounds; see Methods for details). Bold lines indicate significant clusters that survived a permutation analysis. Effects with offset $t < 0$ indicate that reward predicts future proximity/visibility, while effects at $t > 0$ indicate that proximity/visibility predicts future rewards. (**d**) Examples of proximity and visibility networks. (**e**) Average reward as a function of Eigenvector centrality computed on the proximity network. Each dot represents one participant; lines and ribbons show the fixed effect of a hierarchical Bayesian regression (reported above; group rounds in bold). Reliable effects (not overlapping with zero) are underlined. **f**) Correspondence between in- and out-degree of the visibility network. Each dot represents one participant; the regression line is the fixed effect of a hierarchical Bayesian regression (Fig. S4).

proximity predicts high future rewards. Crucially, this pattern is distinct from solo rounds (unimodal positive correlations centered at 0s; see Fig. S3), where we computed the same analysis "as-if" participants were on the same field, allowing us to rule out the role of the environment. Thus, participants closed their social distance when unsuccessful, which then translated into higher future rewards. In turn, higher reward rate predicted reduced proximity, creating a cyclical pattern that was absent in random environments.

Next, we looked at the dynamics of reward and visual field data, where we analyzed the number of visible peers (outbound social attention towards others; Fig. 2b) and the number of observers (inbound social attention; Fig. 2c) at every timepoint. Starting with outbound visibility of peers (Fig. 2b), we found evidence for *adaptive social attention*. Focusing on smooth environments, a negative correlation at offset -2.3s to -1.3s indicates that low reward predicted seeing more peers in the immediate future, while high rewards predicted less visual attention towards others. This indicates adaptive social learning, where poor individual outcomes promote the acquisition of social information. This effect is reversed in the asocial baseline (Fig. S3), providing evidence that it is a distinctly social phenomenon. We also found several negative correlation clusters at positive offsets for both environments, indicating that the opportunity costs for social information acquisition led to lower future rewards. These opportunity cost clusters were different but also present in solo rounds (Fig. S3), suggesting some generic influence of the task structure (i.e., reduced visibility when destroying a block).

Lastly, the dynamics of reward and inbound visibility (number of observers) indicate *success-biased selectivity*, where participants who were the target of social attention acquired higher rewards, both in the past and long into the future (Fig. 2c). In smooth environments, the positive correlation at offset -2.2s to -1.0s indicates that higher rewards predicted more future observers within a short time period, which was absent in random environments and inverted in the asocial baseline (Fig. S3). The strong dip around offset = 0s is due to the visual dynamics of the task, since the splash animation temporarily obscured the avatar when acquiring a reward. However, more notably, we observed two clusters with positive correlations in smooth environments (1.0s to 3.0s and 3.7s to 13.6s), which were absent in random environments and inverted in the asocial baseline. Thus, participants were not only selective in copying individuals who had been successful in the past, they were also able to identify who would be successful in the future (due to the clustered rewards in smooth environments).

## Proximity and visibility networks

We used social network analyses to examine the structure of spatial and visual interactions (Fig. 2d). In *proximity networks* each node is a participant and the undirected edges are weighted by the average proximity (i.e., inverse distance) between players in each round. *Visibility networks* were constructed similarly, but with directed edges weighted proportional to the duration of time that each target player was visible to another player. The same analyses were also applied to solo rounds "as-if" participants had been on the same field.

We first used the proximity network to compute the Eigenvector centrality for each participant. Higher centrality corresponds to participants occupying and maintaining close proximity to others, who in turn are also proximal to others. Whereas centrality tended to correspond to low rewards in random environments (hierarchical Bayesian regression; group: -7.6 [-16.6,0.6]; solo: -7.6 [-16.1,0.8]; all slopes overlap with 0), we found a robust inversion in smooth environments, where centrality predicted higher rewards in smooth environments in group rounds (7.9 [1.6,14.0]; Fig. 2e). This effect disappeared in the asocial baseline (2.6 [-5.6,10.1]), providing evidence that the benefits of spatial centrality were due to social dynamics, since being central afforded more opportunities to imitate others.

Furthermore, we examined the relationship between out- and in-degree in the visibility network. Out-degree is the sum of all outbound edge weights, with more visible peers and longer visibility durations both contributing to larger out-degrees. Similarly, higher in-degree corresponds to more observers with longer durations. This analysis revealed an asymmetry in social attention, with a general inverse relationship between in- and out-degree (Fig. 2f). Whereas this asymmetry was also present in random and solo rounds, it was markedly stronger when combining group rounds and smooth environments (-0.6 [-0.8,-0.4]). This suggests an increased specialization of social learning strategies and asymmetry of social attention in settings where social information was useful (see also Fig. 4d).

## Social influence and leadership

We also found behavioral evidence for changes in social influence based on the frequency of "pull" events (see Methods), characterized by leader-follower dynamics. Inspired by methods used to study collective decision-making in baboons[24], each candidate event was selected from min-max-min sequences in the pairwise distance between players (Fig. 3a) and then filtered by a number of criteria including strength (change in distance relative to absolute distance) and disparity (one player moves more than the other). After filtering, we detected a total of 537 pull events (see Fig. 3a for an example), where in each event, one player is identified as a leader (moved more during $[t_1, t_2]$) and the other as a follower (moved more during $[t_2, t_3]$).

We analyzed both solo and group rounds, with solo rounds providing a benchmark for the sensitivity of these analyses by

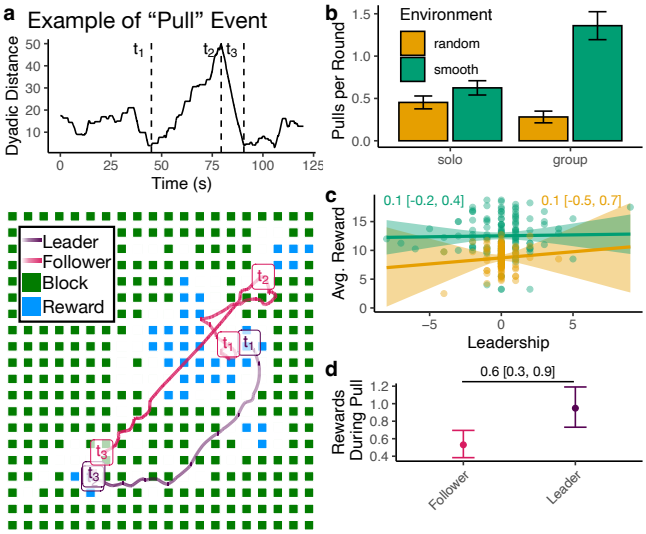


**Figure 3. Social influence**. (**a**) Example of a pull event, selected from min-max-min sequences in dyadic distance and filtered by a number of criteria (see Methods). The trajectories at the bottom are labeled with the three time points that define a pull, and show the state of the environment at time $t_3$. Note that $t_2$ for the leader largely overlaps with $t_3$. (**b**) The average number of pull events per round ($\pm$95% CI). We performed the same analysis on solo rounds "as-if" participants were on the same field to provide an asocial baseline. (**c**) While leadership ($n_{\text{leader}} - n_{\text{follower}}$) did not predict performance, (**d**) leaders had higher instantaneous rewards during pull events.

accounting for the influence of the reward structure (Fig. 3b). While random environments saw a reduction in pull events from solo to group rounds (hierarchical Poisson regression: -0.7 [-1.2, -0.1]), smooth environments saw a large increase in pull events from solo to group rounds (1.4 [0.8, 2.0]). These results were robust to different filter thresholds (Fig. S5) and suggest participants adapted their susceptibility to social influence depending on the relevance of social learning: following others when adaptive (smooth environments), and actively avoiding others when maladaptive (random environments).

Next, we computed a leadership index for each participant based on their frequency of being a leader vs. a follower: $n_{\text{leader}} - n_{\text{follower}}$, using only group rounds for interpretability. Participants with a high leadership index were observed more (i.e., higher in-degree) and observed others less (i.e., lower out-degree), indicating a high correspondence between our analysis of these non-overlapping aspects of the data (i.e., visual field data vs. spatial trajectories; see Fig. S6). Yet neither leadership (Fig. 3c) nor in/out-degree predicted rewards (Fig. S7). However, when we focused on the instantaneous reward rate *during* a pull event (Fig. 3d), we found that leaders received more rewards than did followers (0.6 [0.3,0.9]). Thus, social influence appears to be modulated by success bias. However, we found no long-term benefits of different social learning strategies or leadership.

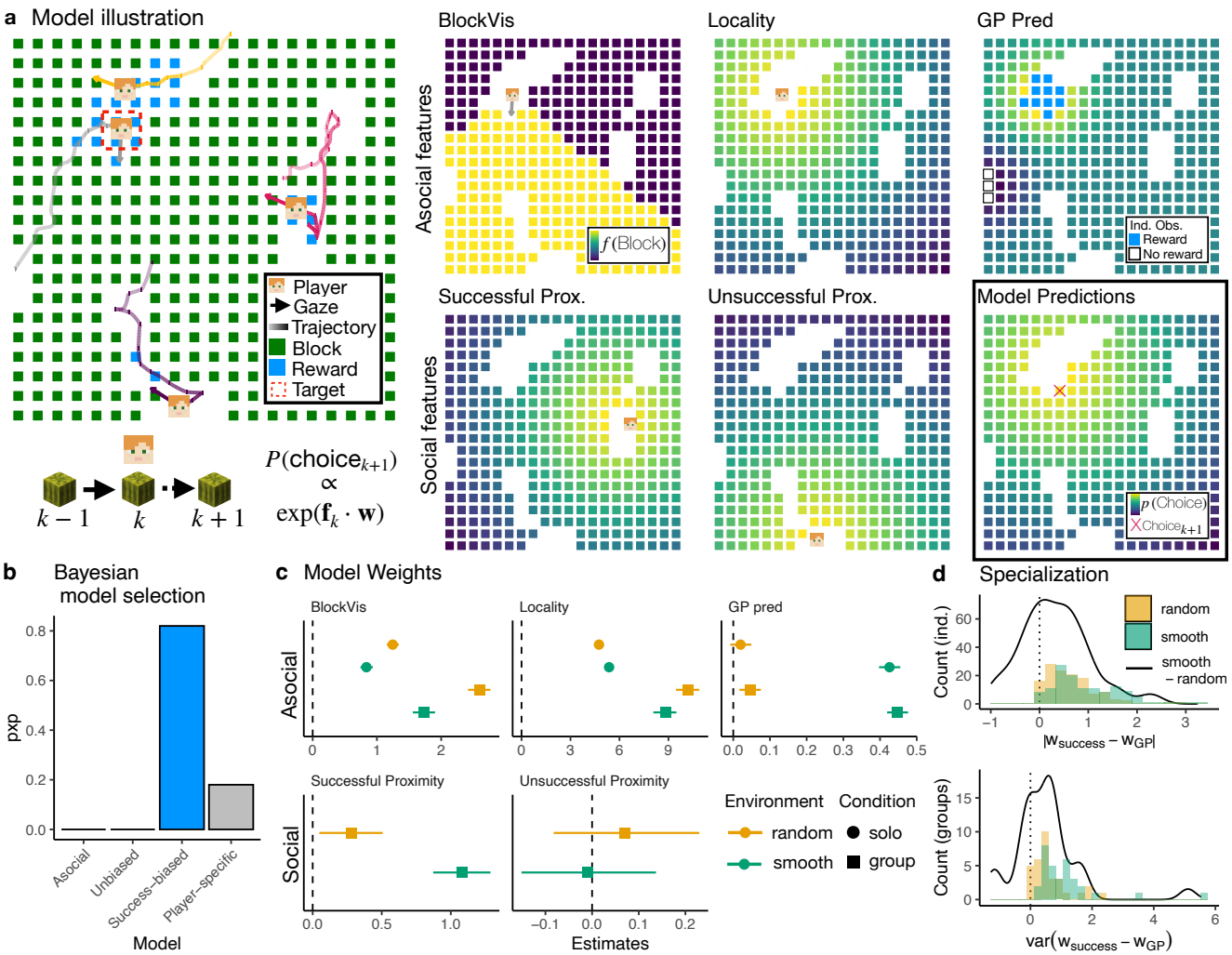


**Figure 4. Computational model**. (**a**) Model illustration focusing on the player highlighted in red. Different models incorporate different sets of features (see text for details), where we illustrate the five features of the winning success-biased model. Increasingly yellow colors correspond to higher feature values. Model predictions (bottom right) are shown using the posterior mean of population-level weight estimates, with the red cross indicating the actual choice. (**b**) Model comparison using protected exceedance probability[25] (pxp) to describe the posterior probability of the best model. (**c**) Population-level weight estimates, showing the posterior mean (dot) and 95% HDI (lines), with a vertical dashed line at 0. (**d**) Increased specialization of learning strategies: individual ($w_{\text{GP}}$) vs. social learning ($w_{\text{success}}$) in smooth vs. random environments, at both individual (top: absolute difference) and group-levels (bottom; variability). Histograms show the raw data, and the density lines show the difference between environments compared to chance (vertical dashed line at 0).

In sum, social learning was highly adaptive (to the environment and depending on individual performance), specialized (asymmetry of social attention), and selective (directed towards successful individuals and with low out-degree). However, different phenotypes of social learning (in/out-degree and leadership index) did not predict overall performance (Fig. 3c; Fig. S7). Only spatial centrality predicted higher rewards (Fig. 2h; Fig. S7f), possibly because close proximity provides more opportunities for adaptive social information use. We next turn to computational models and whether these mechanisms of social proximity and selectivity of social learning predict individual foraging decisions.

## Computational modeling of choices

With a computational modeling framework (Fig. 4a), we sequentially predicted each block participants destroy:

$$P(\text{Choice}_{k+1}) \propto \exp(\mathbf{f}_k \cdot \mathbf{w}) \quad (1)$$

Predictions are modeled as a softmax distribution over a linear combination of block features $\mathbf{f}$ and weights $\mathbf{w}$, where we use the state of the world when the $k$-th block is destroyed in order to predict the $k+1$-th block. Block features $\mathbf{f}$ capture hypotheses about individual and social learning mechanisms (see below), while weights $\mathbf{w}$ are estimated using hierarchical Bayesian methods, controlling for individual and group

variability as random effects (see Methods).

We used a set of asocial features to capture physical constraints of the task and individual learning through reward generalization (Fig. 4a). *BlockVis* captures which blocks are within the player's field of view at time $k$, and is set to 1 if visible and 0 if not (see Methods). *Locality* is the inverse distance to the player at time $k$, reflecting a tendency to forage locally. *GP pred* uses Gaussian process regression as a psychological model of asocial reward generalization, which has been shown to successfully predict human search behavior in a variety of tasks with structured rewards[19,26,27]. Here, we modified the model for binary rewards, where based on the player's reward history (until time $k$), we predicted the probability of each remaining block containing a reward as a logistic sigmoid of a latent variable $z$, with higher values corresponding to higher probability of reward (see Methods).

Additionally, our winning "success-biased" model uses two social features to capture social influence, differentiated by observations of success. *Successful proximity* is computed using players who were visible and were observed acquiring a reward (i.e., visible splash) in the span of $k-1$ to $k$. We used the last observed location of each player to compute proximity (inverse distance), using the centroid if there were multiple successful players. *Unsuccessful proximity* is calculated the same way, but for visible players who were not observed acquiring a reward.

### Model comparison

In addition to the winning success-biased model, we compared several variants (Fig. 4b), with different models defined according to different sets of features. An *asocial* model omitted the social features, while an *unbiased* model did not differentiate between successful and unsuccessful players, combining all visible players into a single unbiased social proximity feature. Lastly, we tested a *player-specific* model, where each individual player had their own hierarchically estimated proximity weight. Using protected exceedence probability ($pxp$) as a hierarchical Bayesian framework for model comparison[25], our success-biased model had the highest probability of being the best model ($pxp(\text{success-biased}) = .82$). The player-specific model was the second best model ($pxp(\text{player-specific}) = .18$), with player-specific weights related to higher rewards, higher in-degree, and higher leadership scores (Fig. S8).

### Model weights

We focus here on interpreting the model weights for the winning success-biased model (Fig. 4c), but all other models had similar weights for shared features (Fig. S9). Block visibility and locality influenced choices in all conditions, and were typically stronger in random than in smooth environments (i.e., in the absence of reward-predictive cues). The one exception is in solo rounds, where participants were more local in smooth environments (0.6 [0.5, 0.7]). GP predictions of reward were strong in smooth environments, and null (solo: 0.02 [-0.01, 0.05]) or negligible (group: 0.05 [0.02, 0.08]) in random environments, corresponding to differences in reward structure.

Social feature weights show that participants were strongly influenced by successful players in smooth environments (1.1 [0.9, 1.3]), with substantially less influence in random environments (smooth - random: -0.8 [-1.0, -0.6]). However, the effect was still reliably different from chance (0.3 [0.04, 0.5]), suggesting a persistence of success-biased imitation even in environments where social learning was irrelevant. In contrast, we found no effect of unsuccessful players in either smooth (-0.01 [-0.15, 0.13]) or random environments (0.07 [-0.08, 0.23]). We also find increased specialization of strategies in smooth environments (Fig. 4d; similar to visual attention asymmetry in Fig. 2f), based on larger differences in social and individual learning weights ($|w_{\text{success}} - w_{\text{GP}}|$: $t(127) = 5.4$, $p < .001$, $d = 0.5$, $BF > 100$), and increased variability at the group-level ($\text{var}(w_{\text{success}} - w_{\text{GP}})$: $t(31) = 2.5$, $p = .018$, $d = 0.4$, $BF = 2.6$).

## Discussion

Using a collective foraging experiment in an immersive virtual environment (Fig. 1), we show how human social learning adapts to different reward structures, is specialized within groups, and selectively directed towards successful individuals. When rewards were smoothly clustered (offering traction for social learning), participants adaptively sought out social information depending on individual performance (Fig. 2a-b), selectively directed their social learning towards successful individuals (Fig. 2c), and specialized more strongly with greater asymmetry of social attention (Fig. 2f). We also detected a greater number of social influence events ("pulls") in smooth environments (Fig. 3b), which were selectively directed towards individuals with higher instantaneous reward rates (Fig. 3d). Our computational models (Fig. 4) combined spatial and visibility data to separately account for both asocial and social learning mechanisms, confirming that success-biased selectivity and specialization are key drivers of adaptive social learning.

The asymmetry of social attention (amplified in smooth environments; Fig. 2f) may act as a safeguard against maladaptive herding[6,7], where instead of copiers copying other copiers, social learning is selectively directed towards individual learners (low out-degree) and is predictive of successful individuals long into the future (Fig. 2c). Although we observed a diverse range of social learning patterns, we failed to find reliable predictors of individual success, with the exception of spatial centrality (Fig. 2e; Fig. S7). This may be due to a combination of social learning strategies having frequency-dependent fitness[28] and the dynamics of selective imitation increasing competition for rewards near successful individuals, leading to equitable performance. While substantially reduced, success-biased social learning was still present in random environments (Fig. 4c), suggesting limitations to the degree of human adaptability and a lingering bias towards social learning, Here, successful foraging outcomes were accompanied by a visual cue, although people can also

deploy metacognitive strategies to infer latent performance from overt behavior[13,29]. Future work can use this paradigm to explore developmental changes[30] or clinical differences in social learning, where our ability to compare all results against an asocial baseline provides a clear demarcation between what is attributable to the environment alone and what is a uniquely social phenomenon.

## Methods

### Participants and design

Participants ($n = 128$) were recruited from the Max Planck Institute for Human Development (MPIB) recruitment pool in Berlin (82 female; $M_{\text{age}} = 27.4$, $SD_{\text{age}} = 5.0$). The study was approved by the Institutional Review Board of the MPIB (number: A 2019-05) and participants signed an informed consent form prior to participation. Participants received a base payment of €12 plus a bonus of €0.03 per reward, spending approximately one hour and earning on average €17.32 $\pm$ 1.02 (*SD*).

Participants completed the task in groups of four. After an in-game tutorial (Supplementary Video 1) and two practice rounds (see below), participants completed 16 2-minute rounds of the task. Using a within-subject design, we manipulated the reward structure (random vs. smooth; Fig. S10) and search condition (solo vs. group). The order of round types was counterbalanced across groups, with four consecutive rounds of the same type (Fig. 1d). The reward structure and search condition for each round was clearly announced prior to the start of each round in an onscreen notification.

The reward structure of a given round was made salient by mapping each reward structure to either pumpkin or watermelon blocks (counterbalanced across groups). In both reward structures, 25% of blocks contained rewards, but rewards were either randomly or smoothly distributed. The smooth environments were generated by sampling from a Gaussian process[31] prior, where we used a radial basis function kernel (Eq. 11) with the lengthscale parameter set to 4 (similar to[19]). Sampled reward functions were then binarized, such that the top quartile (25%) of block locations were set to contain rewards. We generated 20 environments for both smooth and random conditions (Fig. S10), with each session (i.e., group) subsampling 1 (practice) + 8 (main task) = 9 environments of each class with pseudorandom assignments that were pregenerated prior to the experiment. In the tutorial (Fig. 1d), participants were given verbal descriptions of each reward condition, saw two fully revealed illustrations of each environment class from a bird's-eye perspective, and interactively destroyed a 3x3 patch of both smooth and random environments (Supplementary Video 1).

The search conditions were made salient by having participants stand on a teleportation block either by themselves (solo) or with the other three participants (group) in order to begin the round. In the solo condition, participants searched on identical replications of the same environments but without interacting with each other. In the group condition, participants searched on the same environment and could interact with, compete against, and imitate one another.

### Materials and procedure

The experiment was implemented as a computer-based experiment, with each computer connected to a modified Minecraft server (Java edition v.1.12.2). The task was originally designed to also allow for data collection in VR using the Vivecraft mod, which could be done seamlessly by having a participant use a VR headset instead of a mouse and keyboard, without any modifications to the experiment. However, preliminary testing revealed that locomotion via teleportation (the preferred method to avoid VR motion sickness) resulted in less naturalistic spatial trajectories and also interfered with the visual field analyses due to the field of view temporarily fading to black during movement. In contrast, the computer-based modality captured naturalistic trajectories in both space and gaze direction.

In the experiment, the sound was turned off, participants could not see each other's screens, and task-irrelevant controls (e.g., jumping, sprinting, inventory, etc...) were made unavailable. The Minecraft world consists of "blocks" that can be "mined" for resources by holding down the left mouse button to hit them until they are destroyed. In the experiment, participants controlled an avatar that moved through our custom-made environment, defined as a flat field containing a 20x20 grid of 400 pumpkin or watermelon blocks (Fig. 1a) with a two block space between each block. The field was bounded by an impassable fence. See Supplementary Video 2 for a bird's-eye illustration of a round, and Supplementary Videos 3 and 4 for screen captures from group rounds on smooth and random reward environments, respectively.

Each resource block (either watermelon or pumpkin) could be foraged by continually hitting it for 2.25 seconds until it was destroyed, yielding a binary outcome of either reward or no reward. Rewards were indicated by a blue splash effect, visible by other players from any position if it was in their field of view. Only resource blocks could be destroyed in the experiment and destroyed block were not renewed. Blocks did not possess any visual features indicating whether or not they contained a reward. However, rewards in smooth environments were predictable, since observing a reward predicted other rewards nearby. Participants were individually incentivized to collect as many rewards as possible, which were translated into a bonus payment at the end of the experiment. The cumulative number of rewards (reset after the practice rounds) was shown at the bottom of the screen.

After receiving verbal instructions, participants completed an in-game tutorial to familiarize themselves with the controls, how to destroy blocks, the difference between smooth and random reward distributions, and the overall task structure (Supplementary Video 1). They then completed one solo practice round in a smooth environment and one solo practice round in a random environment. These were identical to the solo condition of the main task, but performance in these

rounds did not contribute to a participant's bonus payment. Each round lasted 120 seconds, with the end of the round corresponding to the sun setting below the horizon. This served as an approximate in-game timer for each round, and was communicated to participants in the tutorial. A 3-second countdown timer was also shown onscreen. At the end of the round, participants were given an onscreen announcement indicating the number of rewards they had earned and notifying them of the reward structure and search condition for the next round. Participants were then teleported into a lobby (separate lobbies for solo rounds or a communal one for group rounds), and were required to all stand on a "teleportation" block to indicate readiness for the subsequent round. Prior to the start of a social round, participants all stood on a communal teleportation block, while prior to solo rounds, participants each stood on separate teleportation blocks, in order to induce the social or asocial context. Once all players were ready, a 3-second countdown was displayed and they were teleported into a random position in the next environment with a random orientation direction.

#### *Data collection*

Experimental data was collected using a custom data logging module programmed in Java, which were separated into map logs and player logs. Map logs recorded information about each block destruction event, including a timestamp, player identifier, block position, and the presence or absence of reward. Player logs contained each player's position in the horizontal XZ-plane together with the XYZ components of their gaze vector (i.e., where they were looking). Both logs contained information sampled at Minecraft's native 20 Hz tick-rate (i.e., once every 0.05s), providing high-resolution data about spatial trajectories and gaze directions.

#### *Automated transcription of visual field data*

We developed a custom tool built on the Unity game engine (ver. 2019.3) for performing automated transcription of visual field data (Fig. 1c). We first used data collected from the experiments to simulate each round of the experiment from each participant's point of view. These simulations were then used to automate the transcription of each participant's field of view (Supplementary Video 5).

Our Unity simulations assigned each entity in the experiment (i.e., each block, player, and reward event) a unique RGB value, which was drawn onto a render texture one tenth the size of the player's actual monitor (192x108 pixels as opposed to 1920x1080 pixels). Since the images were rendered without any anti-aliasing or transparency through a simple, unlit color shader, the RGB value of any drawn pixel could be uniquely related with a lookup table to the corresponding entity. We then simulated each round of all experiment data from each player's perspectives within the Unity game engine, using the map logs and player logs, which allowed us to fully reconstruct the world state. Once all four player perspectives were individually rendered, we could read out the pixels from each player's field of view, using the RGB colors of the simulated pixels to determine whether an entity was visible at any point in time (20 Hz resolution), and what proportion of the screen it occupied.

In creating these simulations, a few approximations were required. In addition to the reduced resolution mentioned above, player models were approximated by their directionally-oriented bounding box and we ignored occlusion from the heads up display and view-model (e.g., occlusion due to hand position of the avatar). Additionally, some animations produced by the Minecraft game engine include inherent stochastic processes that were approximated. Namely, the splash particles used to indicate a reward event are generated in Minecraft using a random process that spawns 300 particles at predefined locations in a sphere around the player. Whilst the starting locations are deterministic, small deviations in velocity and the lifetime of these particles are generated randomly. Thus, we tuned the parameters of Unity's particle system to be as authentic as possible by comparing simulated splash effects with video footage of splash effects generated by the Minecraft game engine.

We used a similar procedure for the solo rounds to establish an asocial baseline for our analyses. Whereas all four players searched on different replications of the same field, we simulated them "as if" they were on the same field. Again, a few approximations were required. In these solo simulations, we removed a block whenever any of the four players destroyed it. Additionally, we generated a splash for each reward event, meaning if multiple players foraged the same block in a round, it would trigger a different splash event each time.

#### *Agent-based simulations*

We implemented agent-based simulations to understand how the different reward environments (smooth vs. random distribution of rewards) interact with individual-level learning strategies (asocial learning vs. unbiased imitation vs. biased imitation) in determining foraging success (see Fig. 1f). The simulations uses the same features as the computational model, but are defined fora simplified version of the task, capturing the key visual-spatial dynamics of collective decision-making.

More precisely, our simulations modeled the foraging task as a discrete-time sequential game with partial observability, which generalizes Markov decision processes to incorporate multiple agents, partial observability, and separate rewards[32]. Formally, a task is a tuple, $\langle \mathbb{I}, \mathbb{S}, \mathbb{A}, \mathbb{O}, T, R, O \rangle$: an agent index set, $\mathbb{I}$; a set of environment states corresponding to configurations of agent locations/directions and available/destroyed blocks, $\mathbb{S}$; a set of joint actions corresponding to agents moving in cardinal directions, $\mathbb{A} = \times_i \mathbb{A}_i$; a set of joint observations, $\mathbb{O} = \times_i \mathbb{O}_i$, where each $\mathbb{O}_i$ is a subset of events viewed from agent $i$'s perspective (i.e., other agents' locations, reward reveal events, and available blocks); a environment transition function, $T : \mathbb{S} \times \mathbb{A} \to \mathbb{S}$; a joint reward function $R : \mathbb{S} \times \mathbb{A} \times \mathbb{S} \to \mathbb{R}^{|\mathbb{I}|}$; and a joint observation function, $O : \mathbb{A} \times \mathbb{S} \to \mathbb{O}$.

Agents are modeled as selecting a destination to navigate to, navigating to that destination, and then destroying the

target block (requiring $k = 9$ timesteps in the simulation; approximately equivalent to the 2.25 seconds required to destroy a block and the maximum movement speed of 4.3 blocks/second). Agent policies consist of a high-level controller that transitions among different modes of behavior $n \in \{\texttt{SelectDest}, \texttt{NavTo}(x), \text{ and } \texttt{forage}(k)\}$, where $x$ is a target destination that a low-level navigation controller moves towards and $k$ is a counter for the number of timesteps left to complete the foraging of a block. When in `SelectDest`, the controller samples a destination from $P_{\mathbf{w}}(x) \propto \exp\{\mathbf{f}(x) \cdot \mathbf{w}\}$, where $\mathbf{f} : X \to \mathbb{R}^K$ returns a real-valued feature vector (incorporating both asocial and social mechanisms, the same as in the computational models; see below) for each destination block, and $\mathbf{w} \in \mathbb{R}^K$ are feature weights.

We considered populations of three types of agents. *Asocial agents* used a combination of locality (distance from current location), block visibility (using a 108.5-degree field of view as in the experiment), and asocial reward learning (see the subsection "Gaussian process for binary reward prediction" below). *Unbiased social agents* added an additional feature using the average proximity from observed social partners since the last choice, while *biased social agents* used a similar social proximity feature, but was computed only from social partners that were observed acquiring a reward since the last choice. All feature weights were arbitrarily set to 1. For each of 20 random/smooth environments, we generated 100 simulations for each agent type in groups of four agents (for a total of $20 \times 2 \times 100 \times 3 = 12{,}000$ simulations). Each simulation was run for 400 timesteps. Figure 1f provides the results of the simulations, showing the average total reward collected by agents by environment type (smooth/random) and strategy (asocial/unbiased social/biased social).

#### *Hierarchical Bayesian regressions*

Statistical analyses were conducted using hierarchical Bayesian regressions to simultaneously model the effects of the experimental manipulations (smooth vs. random and solo vs. group), while controlling for random effects of participants and the group they were assigned to. All regression models used Hamiltonian Markov chain Monte Carlo (MCMC) with a No-U-Turn sampler[33] and were implemented using `brms`[34]. For count-based variables (e.g., blocks destroyed or pull events), we used Poisson regression, but report the untransformed regression coefficients for simplicity. All models used generic, weakly informative priors $\sim \mathcal{N}(0,1)$ and all fixed effects also had corresponding random effects following a maximal random-effects procedure[35]. All models were estimated over four chains of 4,000 iterations, with a burn-in period of 1,000 samples.

#### *Temporal dynamics*

Based on methods applied in Neuroscience[36], the temporal dynamic analyses (Fig. 2a-c and Fig. S3) use time-series data from each participant in each round to discover temporal structures in social interactions, where rewards predict future spatial/visual patterns and where spatial/visual patterns predict future rewards.

The time-series variables we used are reward (binary vector), spatial proximity (average inverse distance to all other players), and both the number of visible peers and the number of observers (integer variables $\in$ [0,3]; acquired from the automated transcription of visual field data) at every point in time (20 Hz time resolution). For solo rounds, we computed both spatial proximity and visibility "as-if" participants were on the same field to provide an asocial baseline.

We then computed correlations between each pair of variables `cor(V1,V2)`, where we iteratively time-lagged `V2` from -20 to +20 seconds, with non-overlapping regions of each time series omitted from the data. Each correlation was then $z$-transformed and corrected for chance using a permutation baseline. This chance correction is based on iteratively permuting the order of `V2` and computing the correlation `cor(V1,V2_permuted)` over 100 different permutations (for each correlation). We then subtracted the $z$-transformed mean of the permutation correlations from the target correlation. These permutation corrected correlations are reported as a population-level mean ($\pm$95% CI) in Figure 2a-c and Figure S3.

Lastly, to provide better interpretability of these results, we used a maximum cluster mass statistic[36] to discover temporally continuous clusters of significance at the population level. For each pair of variables $[V1, V2]$ and within each combination of condition (solo vs. group) and environment (random vs. smooth), we used a cluster permutation test to find a threshold for random clusters. This analysis used 10,000 permutations, where for each, we iterated over each individual time series of $z$-transformed (and chance-corrected) correlations, randomly flipping the sign at each time point. We then used a single-sample $t$-test with $\alpha = .05$ to compute which time points (at the population level) were significantly different from 0. This provided a distribution of the duration of temporally continuous clusters of significance in the randomly permuted data. We then used the upper 95% CI of this distribution as a minimum threshold for the actual data, where we applied the same significance testing procedure, but discarded all clusters shorter in duration than the permutation threshold. The surviving clusters are illustrated with bold lines in Figure 2a-c and Figure S3.

#### *Social influence*

We used methods developed to analyze the movement patterns of geotracked baboons in the wild[24] to measure social influence. This allows us to detect discrete "pull" events over arbitrary time scales, where the movement patterns of one participant (leader) pull in another (follower) to imitate and forage in the same vicinity (Fig. 3).

We first computed the pairwise distance between all participants (Fig. 3a) and defined candidate pull events from min-max-min sequences, where we used a noise threshold of 1 block distance to determine what corresponds to minimum and maximum distances. These candidate sequences were then filtered based on strength, disparity, leadership, and

duration in order to be considered a successful pull.

**Strength** $S_{i,j}$ defines the absolute change in dyadic distance relative to absolute distance:

$$S_{i,j} = \frac{|d_{i,j}(t_2) - d_{i,j}(t_1)||d_{i,j}(t_3) - d_{i,j}(t_2)|}{(d_{i,j}(t_1) + d_{i,j}(t_2))\,(d_{i,j}(t_2) + d_{i,j}(t_3))}, \quad (2)$$

where $d_{i,j}(t_k)$ is the dyadic distance between participants $i$ and $j$ at time $k \in [1,2,3]$ (corresponding to the timepoints of the min-max-min sequence). We required pull events to have a minimum strength of $S_{i,j} > .1$, such that they correspond to meaningful changes in spatial proximity rather than minor "jitters" at long distance.

**Disparity** $\delta_{i,j}$ defines the extent to which one participant moves more than the other in each segment, relative to the total distance moved by both participants:

$$\delta_{i,j} = \frac{|\Delta x_i(t_1,t_2) - \Delta x_j(t_1,t_2)||\Delta x_i(t_2,t_3) - \Delta x_j(t_2,t_3)|}{(\Delta x_i(t_1,t_2) + \Delta x_j(t_1,t_2))\,(\Delta x_i(t_2,t_3) + \Delta x_j(t_2,t_3))}, \quad (3)$$

where $\Delta x_i(t_1,t_2)$ is the displacement between $t_1$ and $t_2$. We filtered pull events to have a minimum disparity of $\delta_{i,j} > .1$, such that changes in spatial proximity were asymmetrically driven by one of the interaction partners. Figure S5 shows that our results are robust to changes in the disparity threshold.

**Leadership** is a simple binary filter requiring that the participant who moved more in the first segment ($t_1$ to $t_2$) moved less in the second segment ($t_2$ to $t_3$). We refer to the participant who moved the most in the first segment $\max_{a\in(i,j)} \Delta x_a(t_1,t_2)$ as the *leader* and the participant who moved the most in the second segment $\max_{b\in(i,j)} \Delta x_a(t_2,t_3)$ as the *follower*. Thus, successful pulls are defined as $a \neq b$, where the leader and follower are separate participants.

**Duration** was the final filter, where we required pulls to be at least 3 seconds in duration (since it takes 2.25 seconds to destroy a block). After all filters were applied, the average pull duration was 13.1 seconds $\pm$ 0.09 (SEM).

## Computational modeling

To better understand individual foraging decisions at a micro-level, we developed a computational modeling framework that sequentially predicts each block participants destroy based on different combinations of asocial and social features. We modeled the choice probabilities for each block destruction using a linear combination of block features $\mathbf{f}$ and regression weights $\mathbf{w}$ that represent the influence of each feature for participants' block choices (Eq. 1). This was modeled using a categorical likelihood function with $B_{k+1}$ possible outcomes (i.e., the number of remaining blocks available for choice at time $k+1$), with a softmax link function. Different models incorporate different sets of features in $\mathbf{f}$. Because models only differed with respect to the social features they incorporate and, thus, make identical predictions for solo rounds, the Bayesian model selection (Fig. 4b) was applied to group rounds only.

For interpretability of weight estimates and to allow for identical prior distributions, we $z$-standardized all block features within each choice, with the exception of block visibility, which was coded as a binary indicator. We also omitted the first choice in each round, since most features need to be computed with respect to some previous block destruction. Thus, we only started modeling from the second choice in each round, conditioned on the first choice. Furthermore, while all asocial features were included as predictors for each choice, the social features could be undefined for some choices if the conditions were not met (e.g., no visible players, or no visible and successful players). In these situations, the feature values were effectively set to 0 for all blocks.

All model weights were estimated in a hierarchical Bayesian framework with random effects accounting for differences in the importance of (asocial and social) features among individuals and experimental groups. The models were fit using Stan as a Hamiltonian Monte Carlo engine for Bayesian inference[37], implemented in R v.4.0.3 through `cmdstanr` version 0.3.0.9. We used within-chain parallelization with `reduce_sum` to reduce model run times through parallel evaluation of the likelihood.

To minimize the risk of overfitting the data, we used weakly informative priors for all parameters. We used weakly informative normal priors centered on 0 for all weight parameters, exponential priors for scale parameters and LKJ priors for correlations matrices[38]. To optimize convergence, we implemented the noncentered version of random effects using a Cholesky decomposition of the correlation matrix[39]. Visual inspection of traceplots and rank histograms[40] suggested good model convergence and no other pathological chain behaviors, with convergence confirmed by the Gelman-Rubin criterion[41] $\hat{R} \leqslant 1.01$ . All inferences about weight parameters are based on several hundred effective samples from the posterior[42]. We provide additional details about some model features below.

### *Block visibility*

Since block visibility only captures a static representation of which blocks were visible at time $k$, we computed it with permissive assumptions. Specifically, we assumed no object or player occlusions (i.e., object permanence) and used only the horizontal component of their gaze vector to avoid incorporating noise from vertical jitters. Visibility computations used the true horizontal viewing angle of 108.5 degrees, corresponding to the 16:9 aspect ratio monitors used in the experiment.

### *Gaussian process for binary reward prediction*

Gaussian processes[31] provide a Bayesian function learning framework, which we use as a psychological model of reward generalization for predicting search behavior[19]. Gaussian processes are typically used to learn a function $f : \mathscr{X} \to \mathbb{R}^n$ that maps the input space $\mathscr{X}$ (i.e., the field of destructible blocks) to real-valued scalar outputs such as continuous reward values.

Here, we modify the Gaussian process framework to the binary classification case, where we want to make probabilistic predictions about whether destroying some block $\mathbf{x}$ will yield

a reward $p(r=1|\mathbf{x})$. This can be described as a logistic sigmoid function $S(\cdot)$ of some real-valued latent variable $z$, such that $p(r=1|z) = S(z) = \frac{1}{1+\exp(-z)}$. We set the prior mean $z_0 = \log(\frac{.25}{1-.25})$ such that $p(r=1|z_0) = 0.25$, corresponding to the true prior probability of rewards. Thus, larger values of $z$ correspond to higher-than-chance reward probabilities, while lower values correspond to lower-than-chance reward probabilities.

The latent variable $z$ thus becomes the target of the Gaussian process posterior predictive distribution, computed for some location $\mathbf{x}_* \in \mathscr{X}$ and conditioned on the past set of observations $\mathscr{D}_k = \{\mathbf{X}_k, \mathbf{r}_k\}$:

$$p(r_* = 1|\mathscr{D}_k) = \int p(r_* = 1|z_*)p(z_*|\mathscr{D}_k)dz_* \tag{4}$$

This exact integral in Eq. 4 is analytically intractable, but (i) assuming $p(z_*|\mathscr{D}_k)$ is Gaussian distributed (using the Laplace approximation[31]; see below) and (ii) approximating $p(r_* = 1|z_*) = S(z_*)$ with the inverse probit function[43,44] $\Phi(z_*)$, we obtain a tractable approximation.

We start by defining a posterior on the latent variable $z_*$ corresponding to some unobserved block $\mathbf{x}_*$:

$$p(z_*|\mathscr{D}_k) = \int p(z_*|\mathbf{z}_k)p(\mathbf{z}_k|\mathscr{D}_k)d\mathbf{z}_k \tag{5}$$

The first term $p(z_*|\mathbf{z}_k)$ is a Gaussian distribution that can be obtained using the standard GP posterior predictive distribution[31], while $p(\mathbf{z}_k|\mathscr{D}_k)$ is intractable. However, the Laplace approximation allows us to approximate the latter term using a Gaussian distribution:

$$p(\mathbf{z}_k|\mathscr{D}_k) = \mathscr{N}(\mathbf{z}_k|\hat{\mathbf{z}}_k, (\mathbf{K}+\mathbf{W}+\sigma_z^2\mathbf{I})^{-1}), \tag{6}$$

where $\hat{\mathbf{z}}_k$ is the posterior mode, $\mathbf{W}$ is a diagonal matrix with diagonal elements $S(\hat{\mathbf{z}}_k)(1-S(\hat{\mathbf{z}}_k))$, $\mathbf{K}$ is the $k \times k$ kernel matrix evaluated at each pair of observed inputs (see Eq. 11), $\sigma_\varepsilon^2$ is the noise variance, and $\mathbf{I}$ is the identity matrix. We set $\sigma_\varepsilon^2 = .0001$ as in the environment generating process.The posterior mode $\hat{\mathbf{z}}$ can be obtained iteratively:

$$\mathbf{z}_k^{new} = \mathbf{K}_z(\mathbf{I}+\mathbf{K}_z)^{-1}(\mathbf{r} - S(\hat{\mathbf{z}}) + \mathbf{W}\hat{\mathbf{z}}_k) \tag{7}$$

where $\mathbf{K}_z = \mathbf{K} + \sigma_z^2\mathbf{I}$, $\hat{\mathbf{z}}_k$ is the current estimate of the posterior mode, $\mathbf{z}_k^{new}$ is the new estimate, and $\hat{\mathbf{z}}_k = \mathbf{z}_k^{new}$ at convergence.

Eq. 5 can now be derived analytically as a Gaussian $p(z_*|\mathscr{D}_k) \approx \mathscr{N}(z_*|\mu_{z*}, \sigma_{z*}^2)$, with mean and variance defined as:

$$\mu_{z*} = \mathbf{k}_*^T(\mathbf{r}_k - S(\hat{\mathbf{z}}_k)) \tag{8}$$

$$\sigma_{z*}^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T(\mathbf{W}^{-1} + \mathbf{K} + \sigma_z^2\mathbf{I})^{-1})\mathbf{k}_* \tag{9}$$

where $\mathbf{k}_*$ applies the kernel to the target $\mathbf{x}_*$ and all previously encountered observations $\mathbf{k}_* = [k(\mathbf{x}_1, \mathbf{x}_*), \ldots, k(\mathbf{x}_k, \mathbf{x}_*)]$.

Lastly, we use the inverse probit function $\Phi(z_*)$ as a common method[43,44] for approximating the reward probability as a function of the mean and variance estimates described above:

$$p(r_* = 1|\mathscr{D}_k) \approx S(\mu_{z*}(1 + \frac{\pi\sigma_{z*}^2}{8})^{-1/2}) \tag{10}$$

As a kernel function, we use the radial basis function kernel, which specifies that the correlation between inputs decays smoothly as a function of distance:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{2l^2}\right) \tag{11}$$

The degree of smoothness is controlled by the length scale $l$, which we set to $l = \sqrt{48}$. Note that this is equivalent to the $l = 4$ used to generate the environments, but accounts for the scaling of the coordinate system in the experiment, where each block has an empty tile on each side.

## References


1. Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B. & Tomasello, M. Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science* (2007).
2. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. Identification of the social and cognitive processes underlying human cumulative culture. *Science* (2012).
3. Mesoudi, A., Whiten, A. & Laland, K. N. Towards a unified science of cultural evolution. *Behav. brain sciences* (2006).
4. Henrich, J. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter* (Princeton University press, 2016).
5. Tomasello, M. *The cultural origins of human cognition* (Harvard university press, 2009).
6. Tump, A. N., Pleskac, T. J. & Kurvers, R. H. Wise or mad crowds? The cognitive mechanisms underlying information cascades. *Science Advances* (2020).
7. Toyokawa, W., Whalen, A. & Laland, K. N. Social learning strategies regulate the wisdom and madness of interactive crowds. *Nat. Hum. Behav.* (2019).
8. Morin, O., Jacquet, P. O., Vaesen, K. & Acerbi, A. Social information use and social information waste. *Philos. Transactions Royal Soc. B* (2021).
9. Rahmani, P., Peruani, F. & Romanczuk, P. Flocking in complex environments—attention trade-offs in collective information processing. *PLoS computational biology* (2020).

**10.** Sridhar, V. H. *et al.* The geometry of decision-making in individuals and collectives. *Proc. Natl. Acad. Sci.* (2021).

**11.** Kendal, R. L. *et al.* Social learning strategies: Bridge-Building between fields. *Trends Cogn. Sci.* (2018).

**12.** Enquist, M. & Ghirlanda, S. Evolution of social learning does not explain the origin of human cumulative culture. *J. Theor. Biol.* (2007).

**13.** Hawkins, R. D. *et al.* Flexible social inference facilitates targeted social learning when rewards are not observable. (2022).

**14.** 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 Biol.* (2017).

**15.** Deffner, D., Kleinow, V. & McElreath, R. Dynamic social learning in temporally and spatially variable environments. *Royal Soc. Open science* (2020).

**16.** Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W. & Laland, K. N. The evolutionary basis of human social learning. *Proc. Royal Soc. B: Biol. Sci.* (2012).

**17.** Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J. & Reader, S. M. Individual consistency and flexibility in human social information use. *Proc. Royal Soc. B: Biol. Sci.* (2014).

**18.** Fisher, D. N., Ilany, A., Silk, M. J. & Tregenza, T. Analysing animal social network dynamics: the potential of stochastic actor-oriented models. *J. Animal Ecol.* (2017).

**19.** Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D. & Meder, B. Generalization guides human exploration in vast decision spaces. *Nat. Hum. Behav.* (2018).

**20.** Henrich, J. & McElreath, R. The evolution of cultural evolution. *Evol. Anthropol. Issues, News, Rev. Issues, News, Rev.* (2003).

**21.** Wood, L. A., Kendal, R. L. & Flynn, E. G. Whom do children copy? model-based biases in social learning. *Dev. Rev.* (2013).

**22.** Laland, K. N. Social learning strategies. *Learn. & behavior* (2004).

**23.** Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D. & Couzin, I. D. Exploration versus exploitation in space, mind, and society. *Trends cognitive sciences* (2015).

**24.** Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D. & Crofoot, M. C. Shared decision-making drives collective movement in wild baboons. *Science* (2015).

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

**26.** Wu, C. M., Schulz, E., Garvert, M. M., Meder, B. & Schuck, N. W. Similarities and differences in spatial and non-spatial cognitive maps. *PLOS Comput. Biol.* (2020).

**27.** Wu, C. M., Schulz, E. & Gershman, S. J. Inference and search on graph-structured spaces. *Comput. Brain & Behav.* (2021).

**28.** Rogers, A. R. Does biology constrain culture? *Am. Anthropol.* (1988).

**29.** Wu, C. M., Vélez, N. & Cushman, F. A. Representational exchange in human social learning: Balancing efficiency and flexibility. (Cambridge University Press, Cambridge, 2022).

**30.** Giron, A. P. *et al.* Developmental changes in learning resemble stochastic optimization. (in press).

**31.** Rasmussen, C. E. & Williams, C. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning (MIT Press: Cambridge, MA, 2006).

**32.** Littman, M. L. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, (Elsevier, 1994).

**33.** Hoffman, M. D. & Gelman, A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* (2014).

**34.** Bürkner, P.-C. Advanced bayesian multilevel modeling with the r package brms. *The R J.* (2018).

**35.** Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* (2013).

**36.** Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* (2002).

**37.** Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* (2017).

**38.** Lewandowski, D., Kurowicka, D. & Joe, H. Generating random correlation matrices based on vines and extended onion method. *J. Multivar. Analysis* (2009).

**39.** McElreath, R. *Statistical rethinking: A Bayesian course with examples in R and Stan* (CRC press, 2020).

**40.** Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of mcmc. (2019).

**41.** Gelman, A., Rubin, D. B. *et al.* Inference from iterative simulation using multiple sequences. *Stat. Sci.* (1992).

**42.** Gelman, A. *et al.* *Bayesian data analysis* (Chapman and Hall/CRC, 2013).

**43.** Bishop, C. M. Pattern recognition. *Mach. learning* (2006).

**44.** MacKay, D. J. The evidence framework applied to classification networks. *Neural computation* (1992).

**45.** Wu, C. M. *et al.* Specialization and selective social attention establishes the balance between individual and social learning. In Fitch, T., Lamm, C., Leder, H. &

Teßmar-Raible, K. (eds.) *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*, (2021).

### Data Availability

All data will be made publicly available upon publication.

### Code Availability

All code for running the experiment, analyzing the data, and the Unity simulations will be made publicly available upon publication.

## Acknowledgements

We thank Philip Jakob for technical support, Philipp Schwartenbeck for advice on cluster correction, and Ariana Strandburg-Peshkin and Damien Farine for publishing their code for analysing pull events. We also thank Ryutaro Uchiyama and Alexandra Witt for helpful feedback on a draft of the manuscript, and Deb Ain for copying editing. CMW is 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. RHJMK and DD are funded by the DFG under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135. A pilot version of this experiment (with different data and design) was presented at the 43rd Annual Conference of the Cognitive Science Society[45].

## Author contributions statement

CMW and RHJMK conceived the experiment, with feedback from DD, BK, and BM. CMW, DD, BK, and RHJMK performed the experiments. CMW, DD, and MHH analyzed the results. BK developed the visual field transcription method under the supervision of CMW and RHJMK. CMW developed the visualizations and wrote the first draft. All authors reviewed the manuscript.

## Ethics declarations

### Competing interests

The authors declare no competing interests.

# Supplementary Information for

## The Dynamics of Social Learning in Immersive Environments

**Charley M. Wu, Dominik Deffner, Benjamin Kahl, Björn Meder, Mark H. Ho, & Ralf H.J.M. Kurvers**

## Supplementary Videos

**Supplementary Video 1.** Tutorial. The original German text has been translated to English for better interpretability.

**Supplementary Video 2.** Bird's eye recreation of a group round with smooth rewards.

**Supplementary Video 3.** Screen capture of a group round with smooth rewards (corresponds to Supplementary Video 2). All in-game text was originally in German for all experiments, but have been translated here to English for interpretability.

**Supplementary Video 4.** Screen capture of a group round with random rewards.

**Supplementary Video 5.** Automated transcription of visual field using Unity simulations.

## Supplementary Results

### Rewards, social distance, and foraging rate

Smooth environments resulted in higher rewards in both solo (0.5 [0.4,0.5]) and group conditions (0.3 [0.3, 0.4]; Fig. S1a-c). Within smooth environments, participants performed better in solo than in group rounds (-0.1 [-0.2, -0.1]). However, this effect of solo vs. group condition disappears when we control for faster reward depletion in groups (Fig S1d-f). Thus, reward structure is the key driver of performance.

We then computed the average pairwise distance between participants (Fig S2a-b). Solo rounds provide an asocial baseline by accounting for the influence of reward structure, which we calculated by simulating as if participants were on the same field. This analysis revealed closer foraging proximity in smooth than in random environments (-0.6 [-0.8,-0.4]). Comparing group to solo rounds, participants socially distanced themselves by avoiding one another in random environments (0.3 [0.1, 0.5]), but not in smooth environments (0.0 [-0.2,0.2]). While there was no overall social distancing effect in smooth environments, this analysis may mask more complex dynamics of social influence (see Temporal dynamics).

Lastly, our analysis of foraging rate (i.e., number of destroyed blocks per second) revealed greater selectivity in smooth environments, corresponding to a slower rate of blocks destroyed (-0.04, [-0.07, -0.02]; Fig. S2c-d). The selectivity of smooth environments was further amplified in group rounds (-0.03, [-0.06, -0.005]), where participants did not only need to contend with the structure of the environment, but also the structure and dynamics of social interactions.

In sum, smooth environments increased reward rate (Fig. S1a), brought participants closer together (Fig. S2a), and slowed the rate of foraging (Fig. S2c). Groups performed on par with solo rounds when controlling for depletion (Fig. S1d-f), avoided each other in random environments (Fig. S2a), and foraged even slower in smooth environments (Fig. S2c).

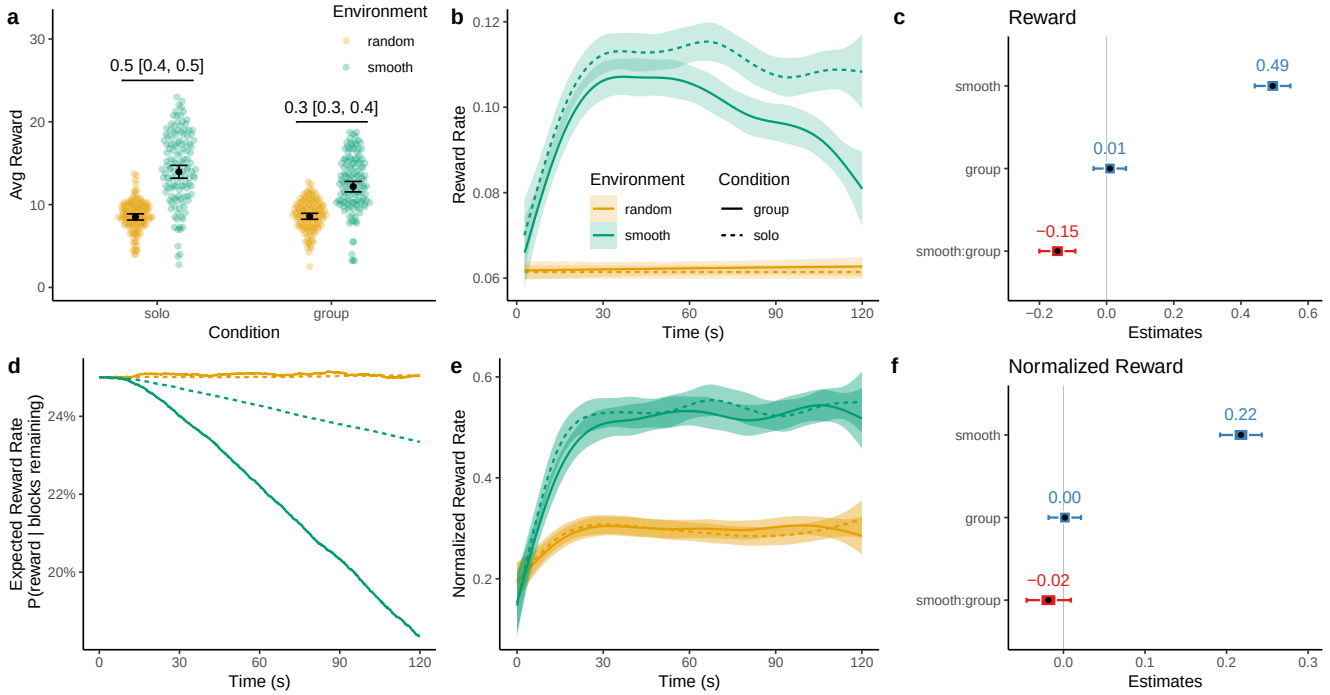


**Figure S1. Reward**. **a**) Average rewards per round. Each colored dot is a participant and the black dot and error bars show the fixed effect ($\pm$ 95% CI) of a hierarchical Bayesian regression (panel c), with contrasts reported above. **b**) Smoothed curves showing the average rate of rewards over time as a Generalized Additive Model (GAM). Ribbons indicate 95% CI. **c**) Coefficient plot of a hierarchical Bayesian Poisson regression showing the rate of rewards. Each dot is the posterior mean and error bars show the 95% CIs. **d**) Expected reward rate over time, showing the probability that a randomly sampled block (from those remaining) contains a reward. Each line shows the aggregated mean, which only diminished in smooth environments (due to predictable rewards) and much faster in group rounds (due to more participants foraging for the same finite number of rewards). **e**) We compute a normalized reward rate by dividing reward rate (panel b) by expected reward rate (panel d). **f**) When running a comparable regression on the normalized rewards, only the effect of smooth environments remains.

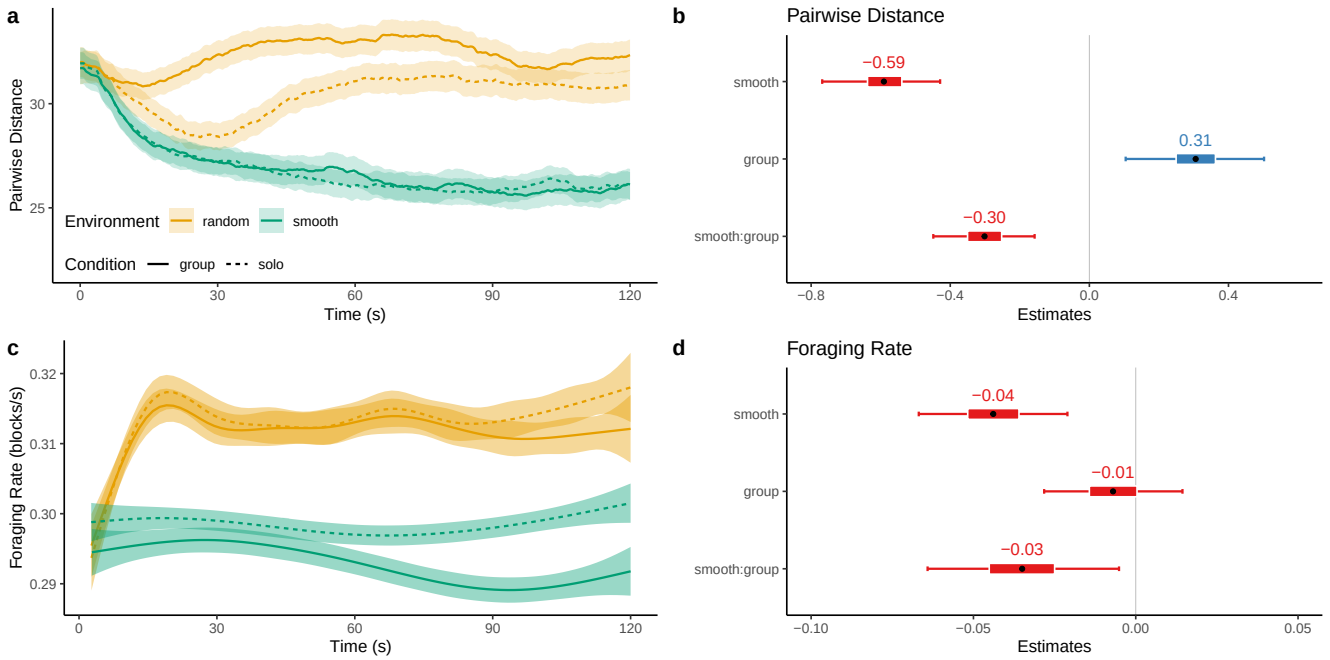


**Figure S2. Pairwise distance and foraging rates**. **a**) Average pairwise distance between players over time, showing the mean (line) and 95% CI (ribbon). **b**) Regression coefficients. Participants were closer together in smooth environments and solo conditions. In random environments, participants in groups avoided each other compared to the solo condition. **c**) Foraging rate (i.e., the number of blocks destroyed per second) plotted as smooth conditional means. **d**) Regression coefficients. Participants had a lower foraging rate in smooth than in random environments, which was further amplified in group rounds.

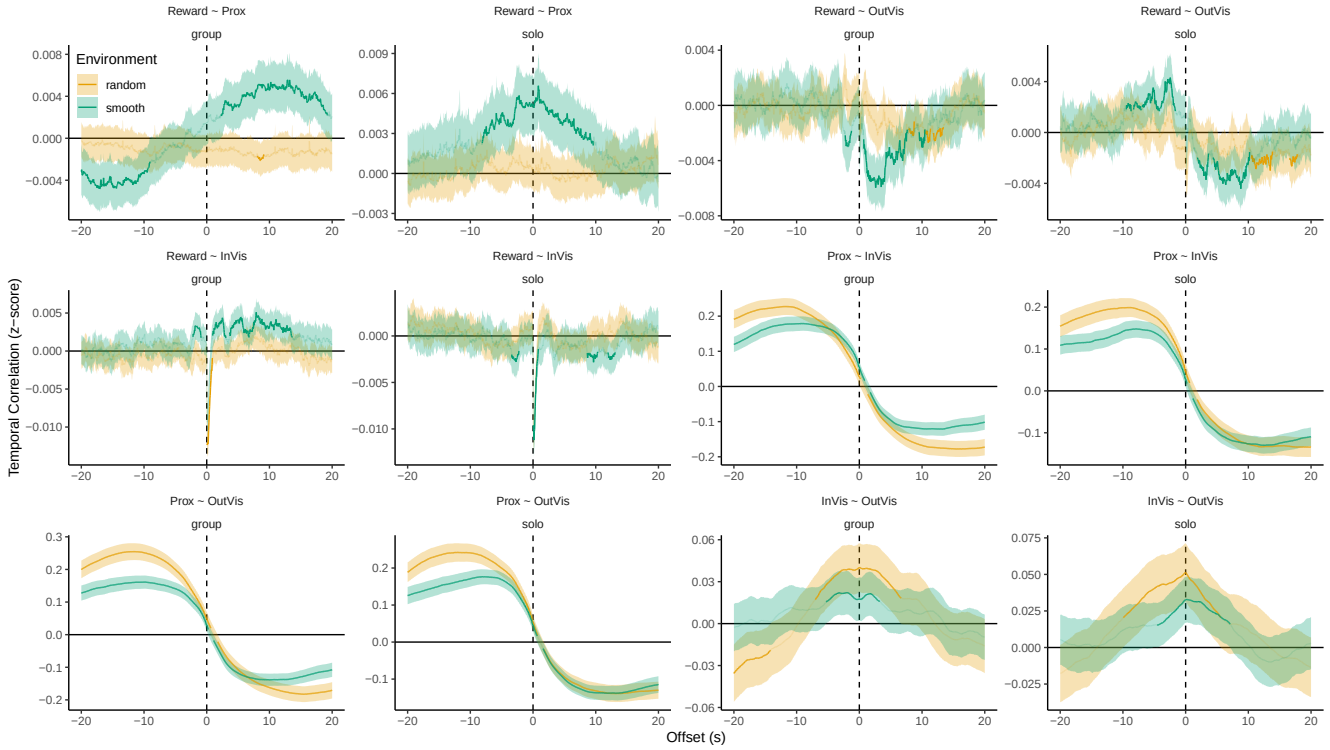


**Figure S3. Temporal dynamics**. Full set of temporal dynamic analyses, including solo rounds. Bold lines indicate significant clusters that survived the permutation analysis (see Methods).

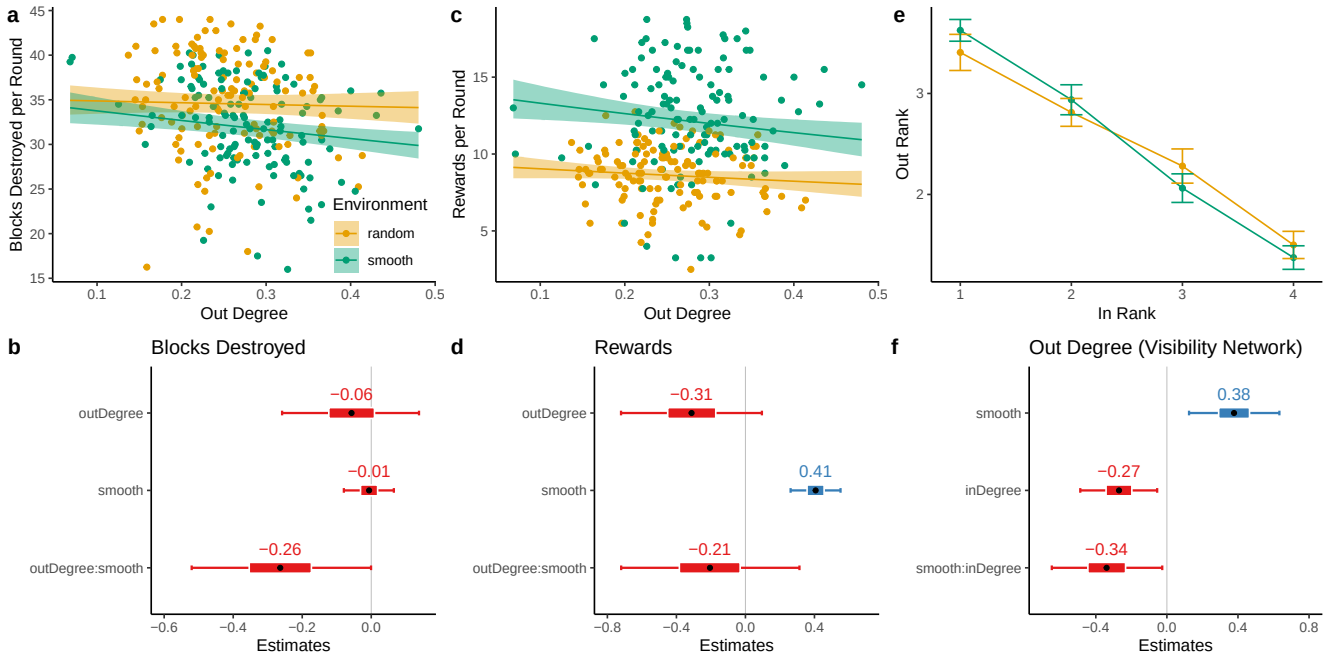


**Figure S4. Relationships between environment, out-degree, in-degree, foraging rate and reward rate in groups**. **a-b**) In smooth environments, participants with a higher out-degree (i.e., observing other players) destroyed fewer blocks. **c-d**) This did not translate into an effect on the reward rate. **e**) Rank ordering participants in each group according to their in- and out-degree, also showed a negative correlation between participants' in- and out-degree. **f**) Regression coefficients predicting out-degree. Smooth environments increased out-degree, while higher in-degree decreased out-degree, with a reliably stronger negative effect in smooth environments.

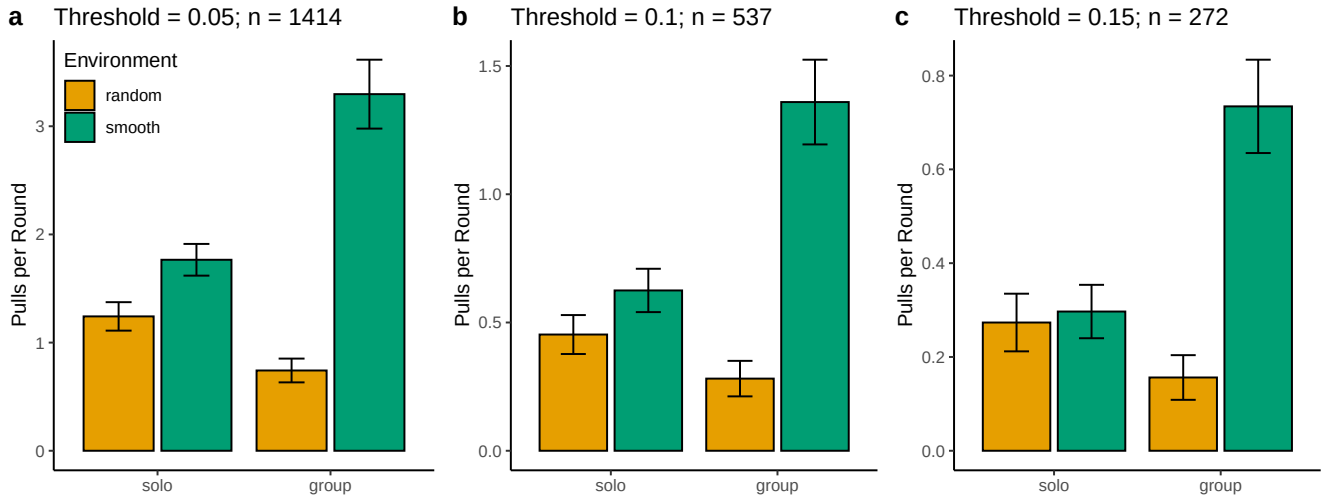


**Figure S5. Sensitivity analysis of the pull analysis. a-c**) Independent of the exact disparity threshold used (0.05, 0.1, 0.15; see Methods), the number of pull events decreased from solo to group rounds in random environments, and increased from solo to group rounds in smooth environments.

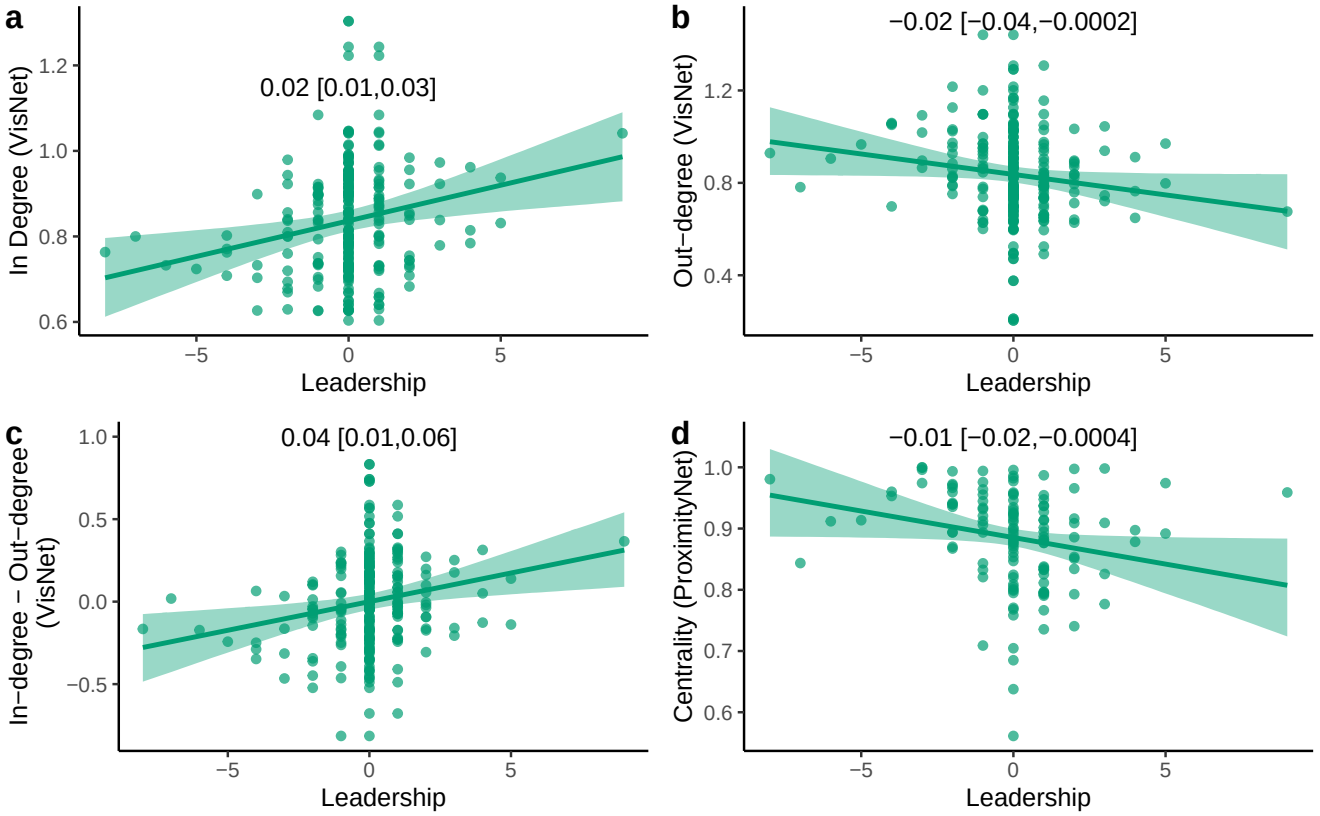


**Figure S6. Relating leadership to visibility (VisNet) and proximity networks (ProximityNet).** Each dot is a participant, with the line and ribbon showing the mean (±95% CI) of a mixed-effects regression, with the fixed effect reported above. For interpretability, we compute leadership only from group rounds in smooth environments. **a**) Participants with a higher leadership score were observed more (i.e., higher in-degree), and **b**) observed others less (i.e., lower out-degree). **c**) Leadership score also predicted the difference between in-/out-degree, and **d**) high leadership score also predicted lower spatial centrality, suggesting leaders were at the frontiers of the group.

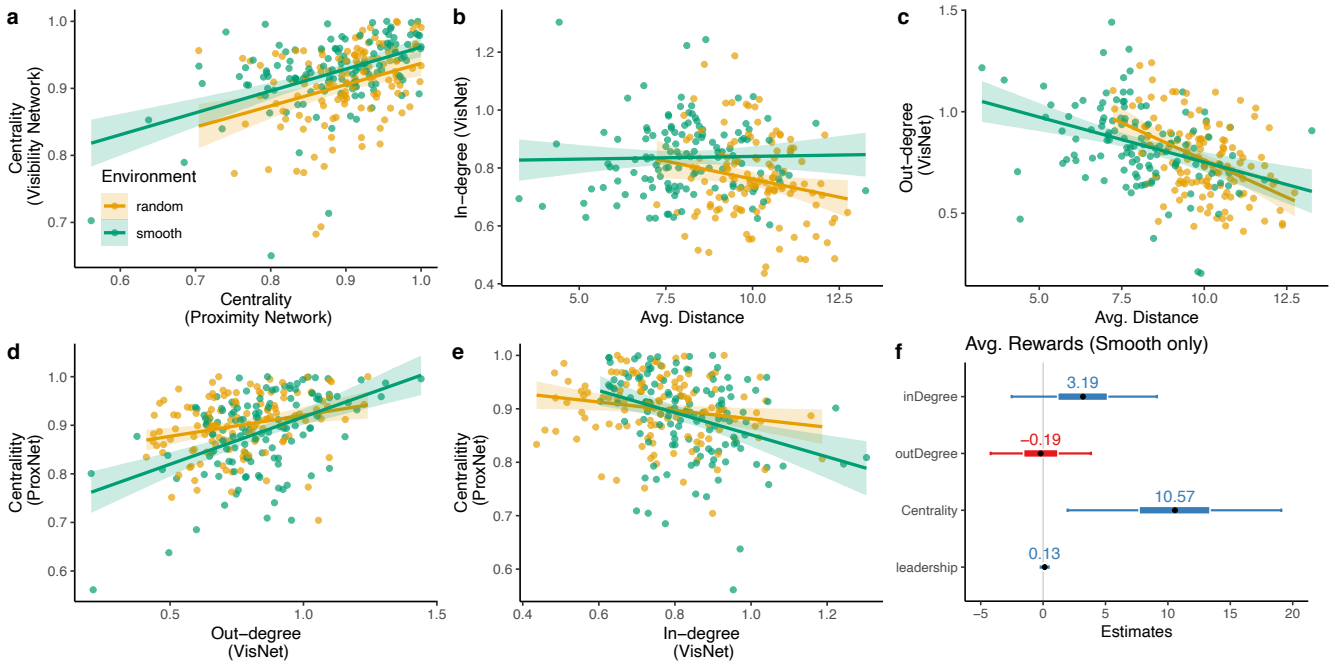


**Figure S7. Consistency of visibility and proximity networks and their relationship to performance (in group rounds).** Each dot is a participant, with the line and ribbon showing a linear regression. **a**) Eigenvector centrality was consistent across visibility and proximity networks. **b**) A player's in-degree was unrelated to their average spatial distance to other players in smooth environments, **c**) **d**) **e**) **f**) Bayesian mixed-effects regression predicting the influence of social network statistics on reward (smooth rounds

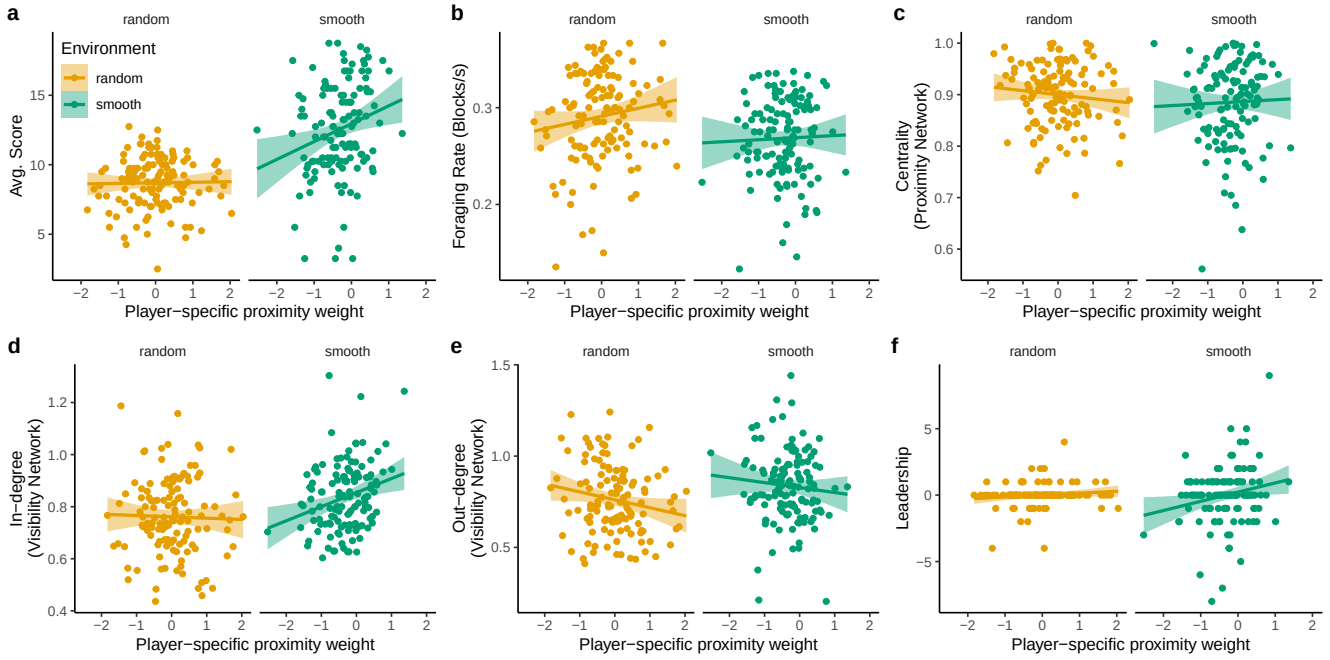


**Figure S8. Player-specific weights**. Higher weights indicate a player exerted more social influence over the rest of their group. Since weights were estimated hierarchically (across the other members of each group) with individual and group as random effects, we report the posterior means. Each dot is a participant, with the line and ribbon showing a linear regression. **a**) Players with higher average score exerted more influence (i.e., larger weights) in smooth ($r = .24$, $p = .007$, $BF = 6.5$), but not in random environments ($r = .02$, $p = .853$, $BF = .21$). **b**) Weights were not related to either foraging rate (random: $r = .14$, $p = .126$, $BF = .63$; smooth: $r = .03$, $p = .726$, $BF = .22$), **c**) or spatial centrality (random: $r = -.10$, $p = .265$, $BF = .37$; smooth: $r = .03$, $p = .739$, $BF = .22$). **d**) Players with higher weights were observed more (i.e., higher in-degree) in smooth ($r = .26$, $p = .003$, $BF = 14$) but not in random environments ($r = -.03$, $p = .760$, $BF = .21$). **e**) Players with higher weights tended to be observed others less in random ($r = -.18$, $p = .045$, $BF = 1.4$), but not in smooth environments ($r = -.09$, $p = .333$, $BF = .32$). **f**) Lastly, higher weights predicted higher leadership scores in smooth ($r = .20$, $p = .023$, $BF = 2.5$), but not random environments ($r = .13$, $p = .152$, $BF = .55$).

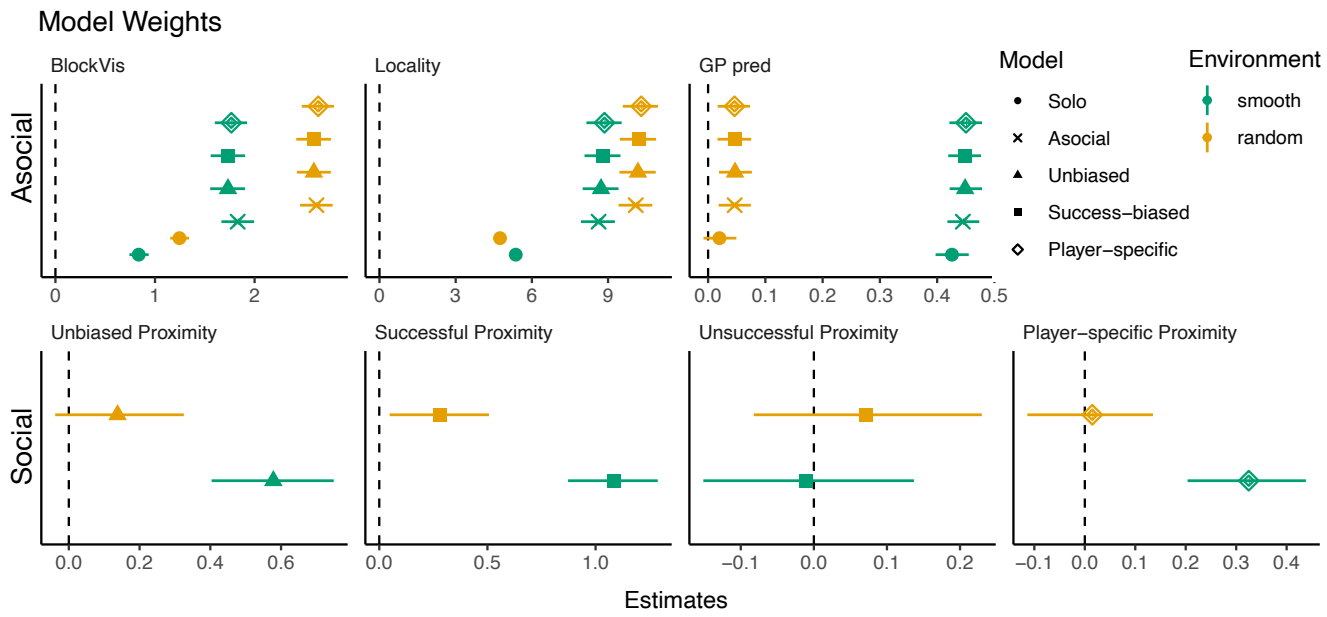


**Figure S9. Weights for all models**. Symbols show the posterior mean and error bars the 95% HDI. The solo model is estimated only on the solo rounds (using only asocial features). The other models use a combination of asocial and social features estimated on group rounds. Player-specific proximity shows the posterior proximity weights aggregated across all players (see Fig. S8 for distribution across individuals).

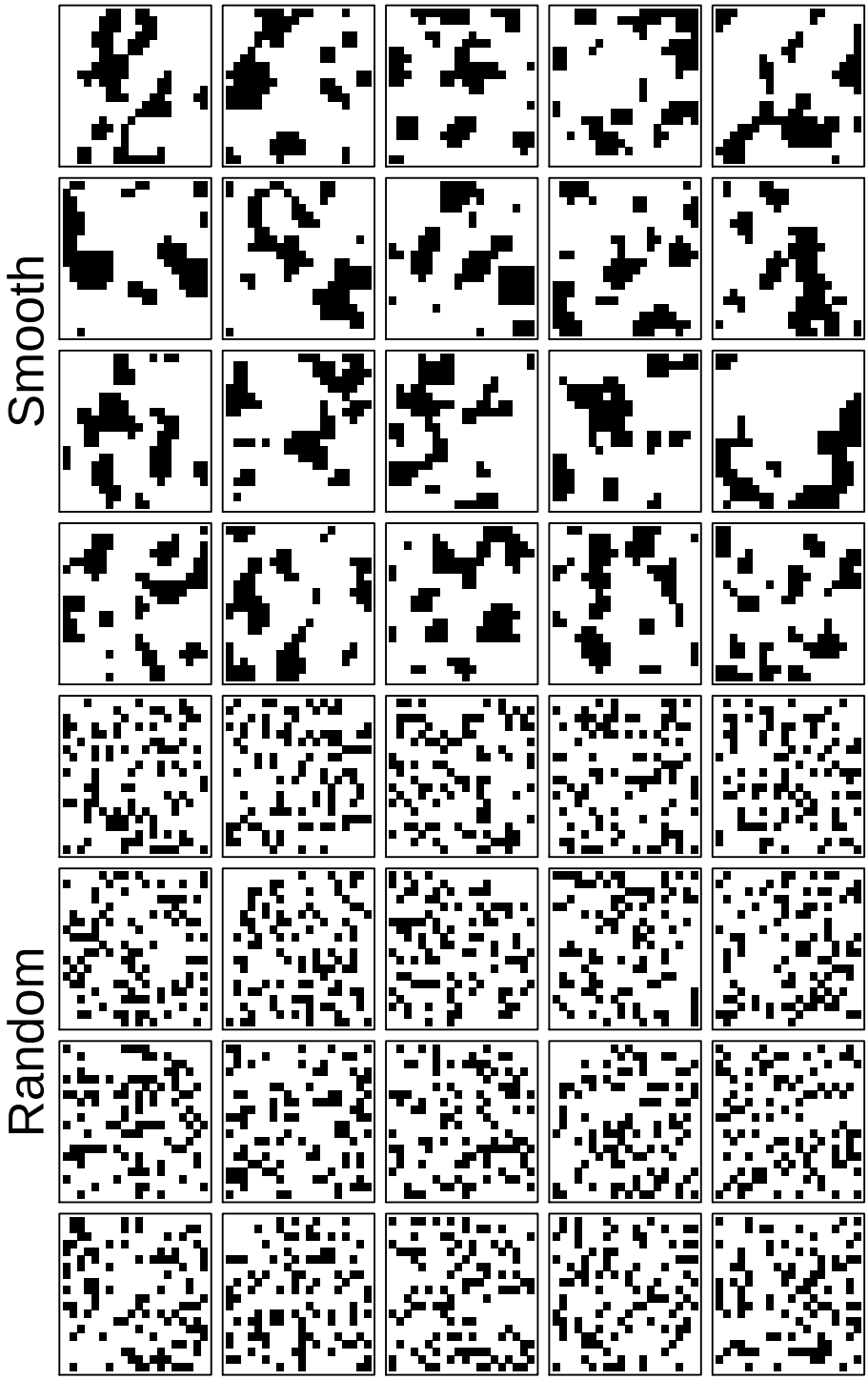


**Figure S10. Reward distributions used in the experiment**. Black squares represent blocks containing a reward, while white squares represent boxes containing no reward. Note that these plots omit the spacing between resource blocks in the experiment for readability.