Article type: Research Article

Title: Globally Inaccurate Stereotypes Can Result from Locally Adaptive Exploration

Authors: Xuechunzi Bai[a,b,1], Susan T. Fiske[a,b], Thomas L. Griffiths[a,c]

Author Affiliations:

[a] Department of Psychology, Princeton University, Princeton, NJ, 08540, US

[b] Princeton School of Public and International Affairs, Princeton, NJ, 08540, US

[c] Department of Computer Science, Princeton University, Princeton, NJ, 08540, US

[1] **To whom correspondence may be addressed:** xb2@princeton.edu

Keywords: social stereotypes, rational analysis, exploration, reinforcement learning

Author Contributions:

T. L. Griffiths and X. Bai developed the research idea. S. T. Fiske, T. L. Griffiths. and X. Bai. designed the experiments. X. Bai. programmed and conducted the simulations, the experiments, and analyzed the data. All authors interpreted the data and wrote the manuscript.

Declaration of Conflicting Interests:

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding:

X. Bai. was supported by the Joint Degree Program in Psychology and Policy, by a Cognitive Science Fellowship, and by the Prize Fellowship in the Social Sciences at Princeton. Additional funding was from T. L. Griffiths. start-up funds. This study is approved by the Princeton Institutional Review Board under protocol number 10859.

Acknowledgments:

We are grateful to Fred Callaway for sharing dynamic programming code, to Jerker Denrell, Joel Martinez, James Walsh, Kevin McKee, and Mark Ho for exchanging stimulating discussions, to Mark Brandt, Klaus Fiedler, and anonymous reviewers for providing constructive reviews.

Supplementary Online Materials:

Additional supporting information can be found in a separate file.

Open Practices:

All data and materials have been made publicly available via the Open Science Framework and can be accessed here. All studies and designs, both pilot and main, have been pre-registered and can be accessed via the same link in SOM.

Disclaimer: This paper was accepted for publication in *Psychological Science*. This is not a final nor copy-edited version of the paper.

Globally Inaccurate Stereotypes Can Result from Locally Adaptive Exploration

Abstract

Inaccurate stereotypes -- perceived differences among groups that do not actually differ -- are prevalent and consequential. Past research explains stereotypes as emerging from a range of factors including motivational biases, cognitive limitations, and information deficits. Considering the minimal forces required to produce inaccurate assumptions about group differences, we show that locally adaptive exploration is sufficient: an initial arbitrary interaction, if rewarding enough, may discourage investigating alternatives that would be equal or better. Historical accidents can snowball into globally inaccurate generalizations, and inaccurate stereotypes can emerge in the absence of real group differences. Using multi-armed bandit models, we show that the mere act of choosing among groups with the goal of maximizing the long-term benefit of interactions is enough to produce inaccurate assessments of different groups. This phenomenon is reproduced in two large online experiments ($N = 2404$), demonstrating a minimal process that suffices to produce biased impressions.

Keywords: social stereotypes, rational analysis, exploration, reinforcement learning

Statement of Relevance

Inaccurate stereotypes about social groups are widespread and consequential, but their origin is puzzling. People often think social groups differ from each other, even absent group-level differences. Here, we demonstrate a minimal process that is sufficient to produce inaccurate stereotypes. Without requiring motivations such as ingroup favoritism, cognitive limitations such as selective attention, or information deficits such as minority representations, a simple, locally adaptive exploratory sampling process is enough to create globally inaccurate stereotypes in an environment with no differences between groups. Our evidence comes from both a formal model and two online experiments. This minimal-process paradigm revisits the origin of inaccurate stereotypes, provides new theoretical insight into this phenomenon, and hypothesizes theory-driven interventions to reduce intergroup misunderstanding.

Introduction

Inaccurate stereotypes about social groups are widespread (Allport, 1954, Ch.6; Fiske & Durante, 2016; Stangor & Schaller, 1996). People believe groups differ from each other, even when they do not.

Why? Explanations for stereotypes' origins fall into three classes (Hilton & von Hippel, 1996; Sherman, Sherman, Percy, & Soderberg, 2013). *Motivational* explanations suggest stereotypes result from humans' priority on belongingness. For example, the minimal-group paradigm merely categorizing people into arbitrary groups causes ingroup favoritism (Tajfel, 1971); social identity theory suggests that stereotypes emerge because people need a positive, distinctive collective ingroup-concept (Brewer, 1999; Tajfel & Turner, 1979). Alternatively, social dominance theory suggests that stereotypes emerge as legitimizing myths that explain the group hierarchy (Sidanius & Pratto, 1999); similarly, system justification theory describes stereotypes as placating, explaining, and maintaining the status quo (Jost & Banaji, 1994). In these explanations, group-serving motivation triumphs over accuracy.

By contrast, *cognitive* explanations suggest stereotypes emerge even without motivational biases. Limited-capacity human minds create shortcuts via schemas (Fiske & Taylor, 1984) and heuristics (Tversky & Kahneman, 1974). For example, categorization makes outgroup members seem interchangeably alike (Taylor et al.,1978). Alternatively, illusory correlation's selective attention to rarity and negativity, given unequal group size, links minorities and negative attributes (Hamilton & Gifford, 1976). Here, cognitive efficiency sacrifices accuracy (Macrae & Bodenhausen, 2000).

Independently, we propose an alternative that considers minimal conditions sufficient to produce inaccurate stereotypes. Mistaken impressions can fall out from maximizing the long-term rewards of interactions: exploring the payoffs from different groups, then letting these rewards guide future interactions. Seeking to maximize long-term rewards in an environment where all groups are equally rewarding suffices to produce inaccurate stereotypes. Our approach is *minimal* because group-serving motivations or cognitive efficiency are not necessary for stereotypes' emergence. Our approach is *functional* because stereotypes emerge as an epiphenomenon of an adaptive solution to maximizing long-term reward. The solution is locally adaptive because it learns only as much about each group as needed to identify one that rewards interactions, without trying to accurately estimate the rewards from each group. This minimal, functional analysis does not imply social stereotypes are accurate or morally right. Rather, it helps explain why stereotypes are so widespread: globally inaccurate impressions can emerge from locally adaptive exploration.

Imagine choosing collaborators from four groups. To work with someone friendly, you need to collect more information. You can ask a small favor from one person each time, then update your group impressions based on their reactions. If more people from one group help, then you might think that group is warmer than other groups. If the goal is interacting with as many friendly people as possible, you might be more likely to interact with people from that group. However, in an environment where each group is equally and highly likely to help, you may never learn that all groups are equally warm. This adaptive strategy settles quickly on a good-enough decision without incurring the costs of prolonged search for other equally good alternatives. As such, a byproduct of maximizing interactions with friendly people in an environment with no group differences is inaccurate stereotypes -- you form accurate impressions about the group with whom you interact most, but form inaccurate impressions about the groups with whom you interact less.

Prior work has examined exploratory sampling. Previous theoretical analyses in social settings have shown that evaluation-based sampling exacerbates inaccurate impressions. For example, the “hot-stove effect” theorizes that people tend to avoid repeating a negative experience, which prioritizes negativity (Denrell, 2005; Denrell & March, 2001). Such a sampling-based approach could explain why people underestimate the trustworthiness of others: If people falsely believe that others cannot be trusted, they avoid them, and by avoiding them, they cannot disconfirm their false belief (Fetchenhauer & Dunning, 2010). An analysis of experience sampling revealed that even an agent with a Bayesian belief-updating process could form biased impressions when information from one group is always available (LeMens & Denrell, 2011). (Person perception, as noted, routinely directs attention and weight to negative information, in the service of avoiding harm (Fiske, 1980; Skowronski & Carlston, 1989). But here we focus on a different phenomenon, seeking reward and avoiding its absence.) Complementing this, non-social settings show initial biases can consolidate in reward-rich environments (Harris, Fiedler, Marien, & Custers, 2020). When two options predominantly and equally yield positive outcomes, the initial bias is upheld because pursuit of the allegedly superior option reinforces the biased preference.

We inherit impression formation as sequential and uncertain, but build on earlier results. First, we make minimal assumptions: Besides minimizing motivations or cognitive limitations, we do not assume differences in the initial bias (Harris, Fiedler, Marien, & Custers, 2020) or information availability (LeMens & Denrell, 2011). We formulate social exploration using multi-armed bandits -- a standard formalism for exploration. We examine the consequences of solving this problem using Thompson sampling (Thompson, 1933), a standard algorithm with optimality guarantees and existing support from human experiments (Gershman, 2018; Schulz, Konstantinidis, & Speekenbrink, 2018). We assume that all social groups reward interaction equally, and the chances of a reward are high. We show inaccurate stereotypes can emerge even in such a minimalist setting (see Theoretical Models). Second, we empirically evaluate the predictions of this model in a social context. Whether people behave according to normative theories becomes especially a puzzle for decisions involving other humans (Hackel, Mende-Siedlecki, & Amodio, 2018; Lockwood, Apps, & Chang, 2020). Two experiments show that participants form inaccurate stereotypes from our minimal-process paradigm (see Empirical Tests).

Theoretical Models

Method

Imagine you face K groups of people, and each group is configured with an unknown probability θ_k of providing a reward in the form of help $r_{t(k)}$, meaning the average help from each group over the long run. The reward at each round is a binary random variable drawn from a Bernoulli distribution, $r_{t(k)} \sim Bern(\theta_k)$, meaning you either receive help or not. The goal is to find the strategy that achieves highest cumulative reward $\sum_{t=1}^{T} r_{t(k)}$, meaning you want to receive as much help as possible. If you knew the group with highest rate of reward k^*, then you would achieve the optimal cumulative reward $\sum_{t=1}^{T} Q_{t(k^*)}$. Maximizing cumulative reward is equivalent to minimizing the expected cumulative regret from not picking the best group,

$$\mathrm{R} = E[\sum_{t=1}^{T} Q_{t(k^*)} - \sum_{t=1}^{T} r_{t(k)}].$$

meaning given that you do not know for sure who will help you, so you will make mistakes occasionally, but you want to minimize your average mistakes (Fig 1).

One solution to this problem is known as Thompson sampling (Thompson, 1933; Agrawal & Goyal, 2012). The idea is to pick a group with probability equal to the probability of it being the optimal choice. The value of θ_k for each group is assumed to be drawn from a Beta distribution, Beta (α, β), with α being the counts of success (e.g., being helped) and β being the counts of failure (e.g., not being helped). The expected reward of each group is α/α+β, meaning your expectation about the group in general will change as you accumulate more experiences with being helped or not. At round t, having observed $S_k(t)$ successes and $F_k(t)$ failures, the algorithm applies Bayesian inference to update the distribution on $\theta_k$to Beta (α+$S_k(t)$, β+$F_k(t)$), meaning you add $S_k(t)$ to success if being helped and you add $F_k(t)$ to failure if not being helped. The algorithm then samples values of θ_k from these distributions, and selects the group for which the sampled value is the largest. This is equivalent to choosing each group with a probability that corresponds to the probability the agent gives to that group having the largest actual value of θ_k. The process iterates[1] (a more detailed description appears in S1.1).

[1] Technically, the problem we consider here has a finite horizon over which we are optimizing, in contrast to the infinite horizon setting where Thompson sampling has been shown to be optimal. The optimal finite horizon solution computed by dynamic programming is also consistent with our results, but we focus on Thompson sampling here based on its previous support in the psychological literature (e.g., Gershman, 2018; Schulz et al., 2018). Further details are provided in S1.1 and S1.4.

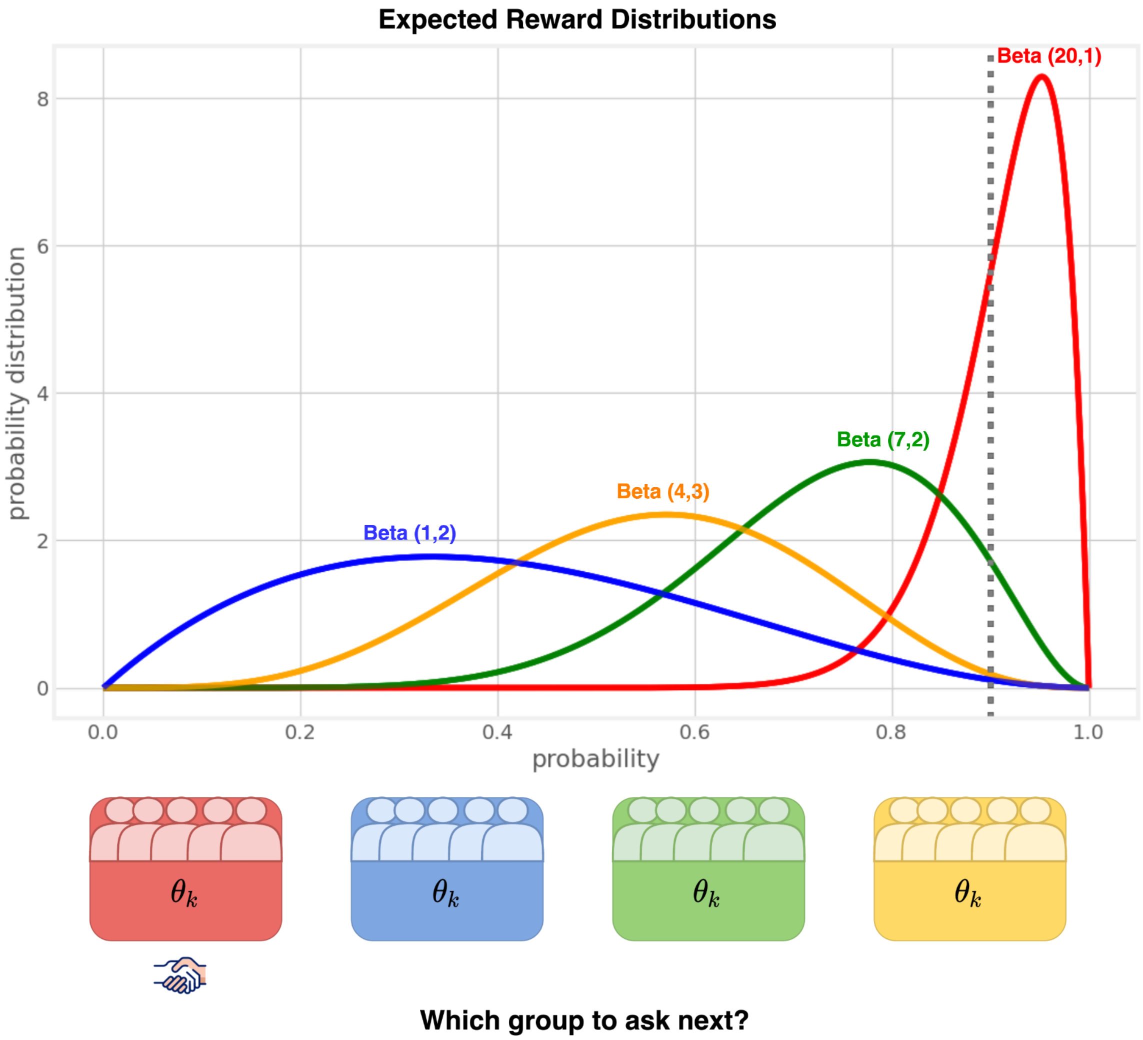


Fig 1. The social multi-armed bandit. *K* groups of people are represented by different colors. The actual reward probabilities (i.e., the dashed-line) are unknown to players. By interacting and observing payoffs at each time, players estimate expected utility for each group. Expected reward distributions parameterized by Beta (α, β) are represented above each group. This example illustrates the agent has 20 success and 1 failure interactions with the reds, 7 success and 2 failure with the greens, 4 success and 3 failure with the yellows, 1 success and 2 failure with the blues. More details in the main text and S1.1.

We explore three main variants of this model that are critical to understanding how inaccurate impressions emerge from locally adaptive exploration. First, we compare sampling strategies: Thompson sampling and random sampling. Thompson sampling proceeds as described above, whereas random sampling selects a group each round according to a uniform distribution. Second, we compare the structure of reward distributions across groups, namely, how the agent behaves when the underlying probabilities of reward are different (e.g., classic bandit setting, Schulz et al., 2018), versus when the underlying probabilities are identical (and reward-rich, Harris et al., 2020). The identical-reward condition specifies $\theta_k = 0.9$ for all k, whereas different-reward condition specifies $\theta_k = \{0.1,\ 0.3,\ 0.5,\ 0.9\}$ for each k. Third, we compare models with and without prior biases. Prior biases reflect to what extent the model expects one group to be better than the others even before collecting any evidence. The prior-bias condition initializes one group with Beta (10,1), meaning the model expects that group to be more rewarding than the remaining groups, whereas the no-prior condition initializes all groups with Beta (1,1), meaning the model expects all groups to be equally rewarding.

Results

We ran 50 simulations, each with 4 groups over a 40-round game (see other simulations with longer time scales, varying ground truth success probabilities, and a dynamic programming algorithm in S1.2-S1.4). Consider each simulation as representing one participant; each participant has 40 chances to select people from one of the four groups sequentially. Given that each simulation has its own most to least selected groups, we rank ordered the results per simulation to make them comparable. Two critical outputs are examined: What is the total number of interactions with each group? What is the estimated expected utility for interacting with each group? Number of interactions can be considered as a behavioral antecedent of impressions, whereas estimated expected utility can be considered as impressions. Hence, [dis]similarity of estimated expected utility among groups against the ground truth utility can be considered as [in]accurate stereotypes.

Our results show that the Thompson sampling model in the identical-reward condition without prior bias, rather than interacting equally and estimating equal rewards for all groups, selectively interacts with one group and estimates that group has higher expected utility than other groups (Fig 2a-b).

To understand why this is, consider a simulated agent with no prior bias, starting the game giving each of the four groups a distribution of Beta (1,1). This assigns equal probability to all values between 0 and 1, so all groups have an expected utility of .5 at the initial round. The agent first samples values from the four distributions and selects the group that has the largest sampled value. Assume the selected group is red, as in Fig 1, so the agent interacts with one member of the red group and observes its reward. If successful, it updates the distribution of the red group to Beta (2,1), yielding an expected utility of .67 while other groups' expected utility remains unchanged. Repeating this process, the agent will sample values from the four distributions again and select the group that yields the largest sampled value. Given that the red group has a higher expected utility than other groups, it is more likely to yield a larger sampled value, thus be selected again. Say the agent selects the red group again. If the interaction is successful, the distribution of the red group will be updated to Beta (3,1) with an expected utility of .75. If unsuccessful, the distribution will be updated to Beta (2,2) with an expected utility of .5. The process iterates. Our particular social environment assumes that all groups have identical and high rewards, making successful interactions more probable. As a result, the agent will be less likely to explore the other three groups thus be less likely to update their expected utility as needed. Inaccurate impressions about the under-explored groups thus emerged in the identical and high-reward environment, purely as a result of locally adaptive exploration.

To compare, first, the Thompson sampling model performs as expected in the different-reward condition. It interacts more with the group that has the highest expected utility, and estimates expected utility for most groups accurately (Fig 2c-d). Second, the random sampling model in the identical-reward condition interacts equally with all groups and estimates expected utility for all groups more accurately (Fig 2e-f). Third, prior biases make the Thompson sampling model converge faster to the ostensibly best group (Fig 2g-h).

Our simulation results show how, without any prior biases, motivational biases, cognitive limitations, or information deficits, agents engaging in a locally adaptive exploration process with the goal of maximizing long-term rewards will form inaccurate impressions, estimating one group as being better than other groups despite the fact that all groups are equally good. We now examine whether the same phenomenon occurs for human participants.

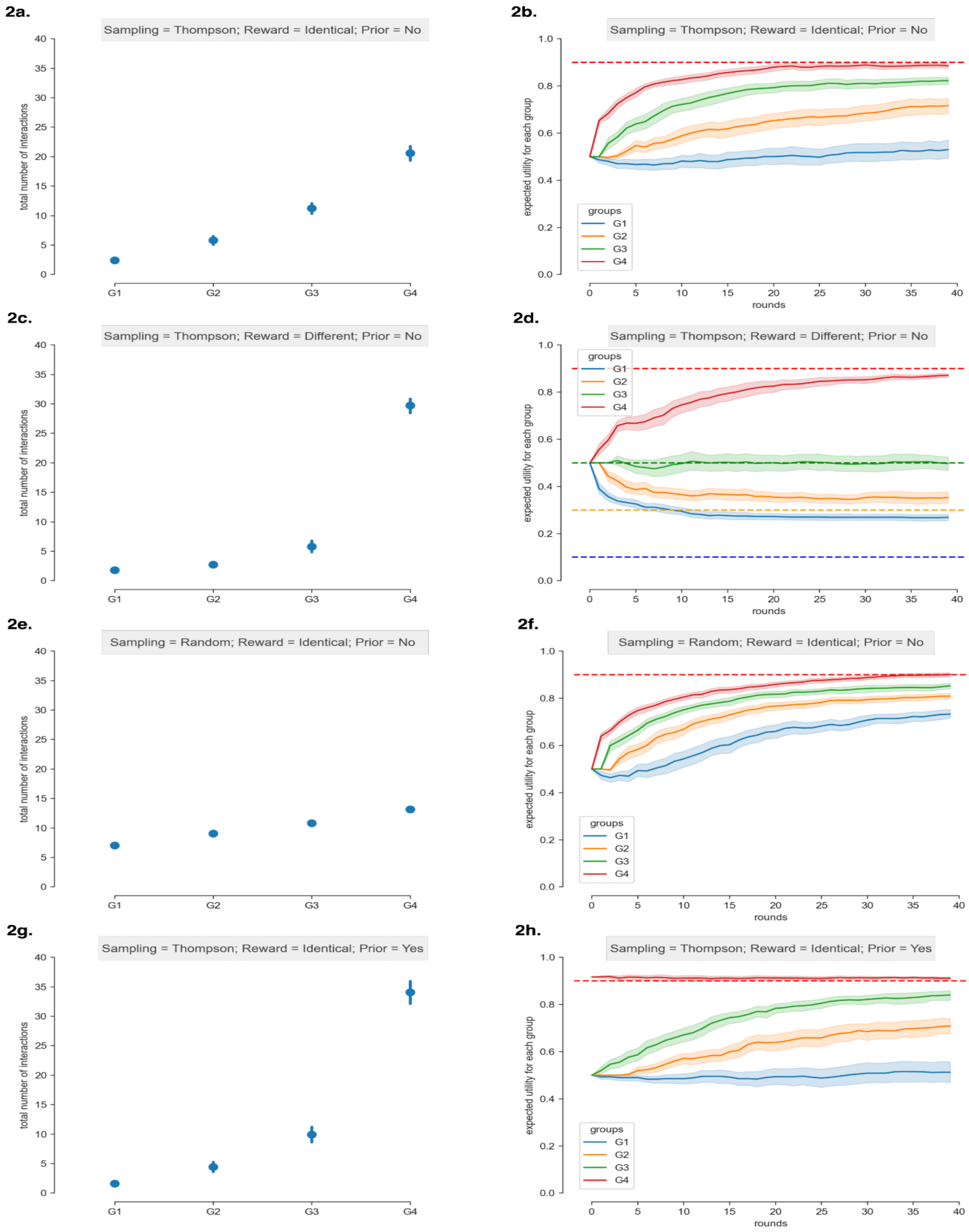


Fig 2. Social multi-armed bandit simulations. Figures display frequency of interactions (point estimates with 95% bootstrapped confidence intervals) and expected utility (point estimates connected by lines over rounds with 95% bootstrapped confidence intervals) for each of the four groups, ordered by frequency of interaction prior to averaging (G4 indicates the most selected group in each simulation). Dashed lines in the right column represent true utility. See interpretations in the main text and more simulations in S1.2-S1.4.

Empirical Tests: Experiment 1

Method

To test the predictions of our theoretical models, we created a narrative travel game called "Explore Toma City" to simulate how people form impressions through social interactions. Participants are invited to a fictional city where they meet people from four novel social groups: Tufa, Aima, Reku, and Weki. Participants learn about these people by interacting with them through 40 rounds of games. Participants can increase the points they earn by starting a small business in the city. They can select people to help them. Some people will increase and share the earned points with the participants, but some will not. In each round, participants get to choose one person, with a new set of people to choose from each time. If the person selected helps to grow the business, then participants earn 1 point (equivalent to a 1 cent monetary bonus at the end of the experiment). If not, participants earn 0 points. Participants can see their points after each decision. After completing the game, participants are asked to estimate rewards ("for each group, how many times out of 100 do you think working with a person from that group would result in you earning 1 point?"), perceived warmth, and perceived competence ("on a scale from 1 to 5, rate how warm/competent the group is"). The fictional journey ends with participants giving suggestions to their friends who are going to visit Toma City in the future (see experiment designs and demo in S2.1-2.2).

To assess participant decisions, we construct two dependent measures (see also Bai, Ramos, & Fiske, 2020): partner choice Herfindahl score and reward estimation standard deviation. These two measures also correspond to the [un]equal number of interactions and [dis]similarity in expected utility from our theoretical model. Unequal number of interactions indicates selective partner choices and dissimilarity in expected utility in a reward-identical condition indicates inaccurate stereotypes. Partner choice Herfindahl score is defined as

$$1 - \sum_{k=1}^{K} S_k^2.$$

where S_k is the observed share of interactions with group k in Toma City. A low score indicates participants predominantly interact with one group, whereas a high score indicates participants approach each group more or less equally. Reward estimation standard deviation is defined as

$$\left(\sum_{k=1}^{K} (x_k - \mu)^2\right)^{1/2}.$$

where x_k is the estimated reward for each group and μ is the average estimation for all groups. A high score indicates participants think groups are very different from each other, whereas a low score indicates participants think groups are more or less similar.

Within this paradigm, Experiment 1 manipulates (a) the sampling strategy: In the Thompson-sampling condition, participants are asked to make decisions themselves (self-select: "Select one person to help you"); we predict participants naturally use this strategy. In the random-sampling condition, participants are shown preselected choices (random-meet: "Meet one person to help you"), in which the choices are randomly assigned by the program; (b) the structure of the reward distribution: In the different-reward condition, the underlying reward distribution is pre-programmed as .1, .3, .5, and .9 for Wekis, Aimas, Tufas, and Rekus respectively. In the identical-reward condition, the underlying reward distribution is pre-programmed as .9 for all groups; (c) the existence of prior bias: In the no-prior condition, participants start the game immediately after entering the city. In the prior-bias condition, before the game, participants see stereotype information associating one group as more competent and warmer than others (e.g., “Rekus are wealthy and very generous to their neighbors”). The game and parameters are programmed on Qualtrics via JavaScript (see materials in S2.1).

Our primary hypothesis predicts participants in the identical-reward, no-prior bias condition with self-selected strategies to be more likely to interact selectively with one group than others (i.e., lower partner choice Herfindahl score) and to be more likely to estimate groups to have different expected rewards (i.e., larger reward estimation standard deviation), than participants in the identical-reward, no-prior bias with random-meet strategies. The hypothesis would fail if we do not observe a statistically significant difference between the two conditions.

Participants

Following a power analysis with pilot data (see pilot details in S3.1), we recruited 399 online workers via Amazon Mechanical Turk’s Cloud Research high-quality pool with 50 participants in each of the 8 conditions, in order to detect medium-to-large effects. Each participant was randomly assigned to one of the 2 by 2 by 2 conditions: identical or different reward, self-select or random-meet strategies, and prior-bias or no-prior information.

Results

As predicted, a simple linear regression with condition as the independent variable (random-meet coded as 0 v. self-select coded as 1) and the Herfindahl score as the dependent variable, participants in the identical-reward and no-prior with self-select condition, as compared to those in the random-meet condition, were more likely to show a lower Herfindahl score (*b* = -.226, 95% *CI* [-.306, -.146], *p* < .001; μ(random-meet) = .75, δ(random-meet) = 0, *N* (random-meet) = 45, μ(self-select) = .52, δ(self-select) = .27, *N* (self-select) = 55, Cohen’s *d* = 1.15), indicating they were more likely to interact with one group than other groups. To situate this comparison, across 40 interactions, participants in the random-meet condition interacted 10 times with each group. However, participants in the self-select condition interacted with the perceived best group on average 22 times, with the second, third, and worst perceived group on average 10, 7, and 5 times (Fig 3a; see individual plots in S4.1-2).

Also as predicted, a simple linear regression with condition as the independent variable (random-meet coded as 0 v. self-select coded as 1) and the standard deviation of the estimated rewards as the dependent variable, participants in the self-select condition were also more likely to show a larger standard deviation of the estimated rewards of the four groups than participants in the random-meet condition ($b = 11.354$, 95% *CI* [7.387, 15.321], $p < .001$; μ(random-meet) = 6.60, δ(random-meet) = 5.01, *N* (random-meet) = 45, μ(self-select) = 17.95, δ(self-select) = 12.61, *N* (self-select) = 55, Cohen's $d = 1.14$), indicating they perceived the groups to differ from each other. To situate this comparison, participants in the random-meet condition estimated the rewards to be on average 93, 88, 83, and 76 (out of 100) points. In contrast, participants in the self-select condition estimated the rewards to be 86, 67, 57, and 42 points (Fig 3b).

An individual-level analysis reveals that the more a participant interacted with one group predominantly, the more likely that participant reported a larger standard deviation in reward estimations (Pearson's $r(98) = -.605$, $p < .001$). This is consistent with previous real-world findings that, the less diverse samples people see, the more distinct stereotypes they have (Bai, Ramos, & Fiske, 2020).

Secondary analyses first confirmed that the structure of the reward distribution matters. Participants in different-reward conditions indeed made good choices to interact with the best group and estimated more accurately on the underlying rewards (Fig 3c-d). Within the context of our design, inaccurate impressions were not absent but less likely to occur when the underlying rewards were different than identical. Next, prior biases matter. In identical-reward conditions, participants with prior biases reported an even larger standard deviation in reward estimations than participants with no priors. Existing stereotypes can produce inaccurate impressions (Fig 3e-f), although existing biases did not have significantly more influence in different-reward conditions than identical-reward conditions (Fig 3g-h) (see analysis details in S4.1-3a for partner choices and S4.1-3b for reward estimates).

Exploratory analyses examined perceptions on warmth and competence of Toma groups (Fiske, Cuddy, Glick, & Xu, 2002). Because succeeding in business requires both warmth and competence, one may expect estimated rewards to correlate with perceived warmth and competence. This intuition is confirmed ($r(397) = .566$, $p < .001$). Similar to estimated rewards standard deviations but with smaller effect sizes (see analysis details in S4.1-4): Participants in the identical-reward condition who made their own choices to interact with Toma people tended to perceive Toma groups as more different to each other than those who randomly met Toma people ($b = .768$, 95% *CI* [.366, 1.170], $p < .001$; μ(random-meet) = 1.50, δ(random-meet) = 0.87, *N* (random-meet) = 45, μ(self-select) =2.27, δ(self-select) = 1.11, *N* (self-select) = 55, Cohen's $d = .77$).

In sum, confirming the theoretical model predictions, human participants exploring the Toma City, with no prior stereotypes, perceive significant differences between Tufas, Aimas, Rekus, and Wekis, when in reality there are no group-level differences. Their behaviors are consistent with the pattern of exploration produced by the Thompson sampling model. Although unintended, inaccurate impressions resulted from locally adaptive exploration.

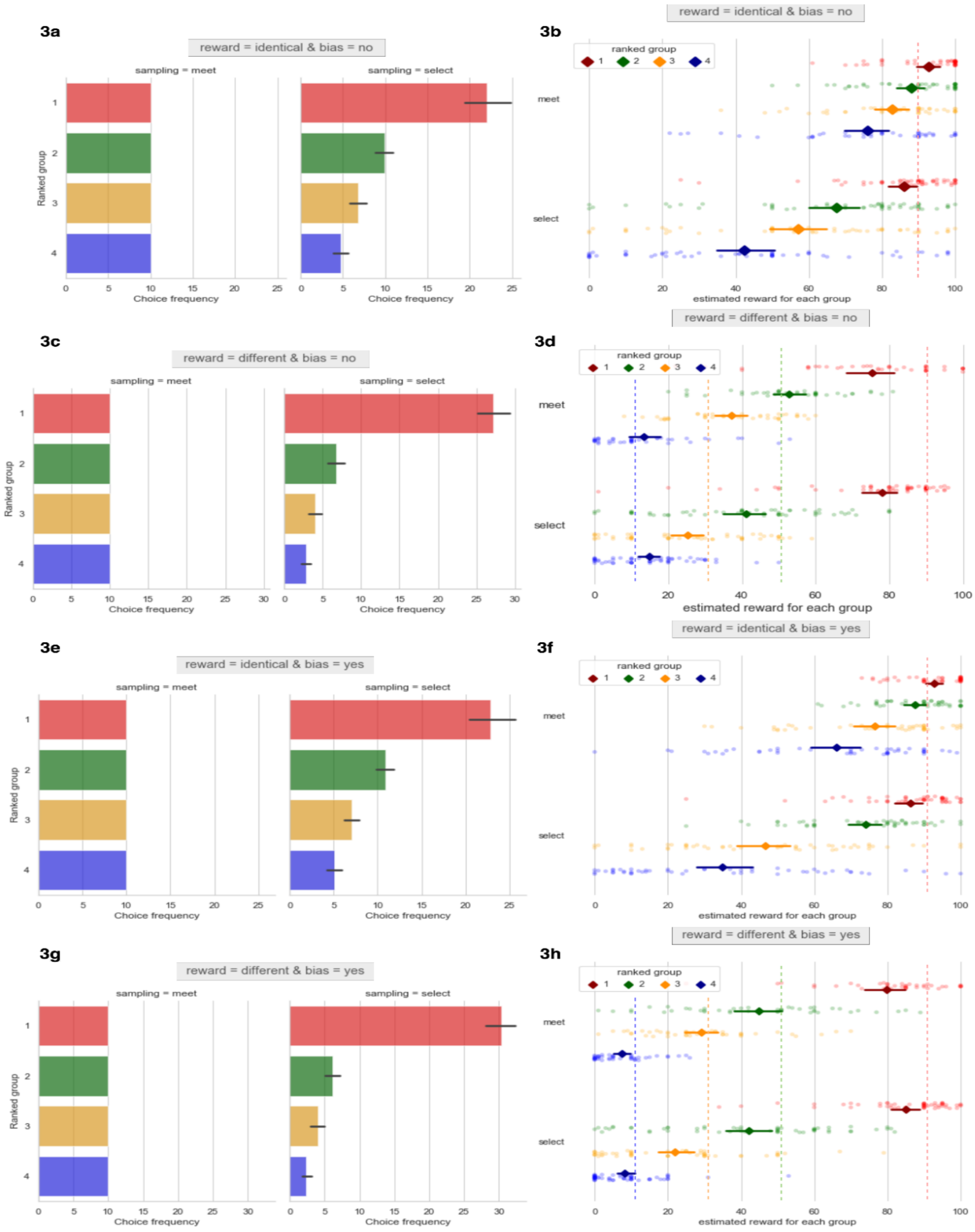


Fig 3. Explore Toma City Experiment 1 showing main predictions. Plots display empirical results of numbers of interactions (bar plots with 95% bootstrapped confidence intervals) and estimated rewards in the final round (point estimates with 95% bootstrapped confidence intervals with raw data in the background), for each of the four groups, ordered as in Figure 2. Dashed lines in the right column represent true utility. See interpretation in the main text and more analyses in S4.1.

Empirical Tests: Experiment 2

Method

Despite statistically significant and large effect-size differences between the self-select and the random-meet conditions in Experiment 1, the precise mechanism is unclear. Experiment 2 investigates two possible mechanisms behind the differences between the self-select and the random-meet conditions: (a) active versus passive learning, and (b) exposure to choices versus rewards (see pre-registration in S3.2).

First, we explored the role of active versus passive learning. Do the selective samples make participants biased, or does the sense of control make them biased? To address this confound, we added a between-subject yoked design. In the yoke-both condition, participants do not have a chance to select partners but can only view partners selected for them (like the random-meet). However, the choices were not randomly drawn from a uniform distribution (unlike the random-meet) but were paired from other participants' responses from the self-select condition (see also Markant & Gureckis, 2014; Prager, Krueger, & Fiedler, 2018). Our theoretical model would not predict a difference between the two conditions, but active learning theories (Bruner, 1961) suggest a difference.

Second, we explored the role of choice exposure versus reward exposure. Are the rewards attached to each choice important for the estimations, or is the mere presence of choices sufficient for biased impressions? To address this, we added another between-subject yoked design. Participants could either encounter the choices and the rewards in the exact sequence (i.e., yoke-both), or they could encounter the choices in the same order but the rewards in randomized order (i.e., yoke-choice-only). Our model would predict that reward order matters, given the sequential feedback of this decision process. Any perturbation of the reward order should then lead participants to form less biased impressions than the exact order (see model details in S3.2), but the mere exposure hypothesis suggests little difference (Zajonc, 1968).

Therefore, Experiment 2 tests four sampling strategies: (a) self-select, (b) yoke-both, (c) yoke-choice-only, and (d) random-meet with the same game using identical-reward and no-prior designs. We predict participants in the self-select, yoke-both, and yoke-choice-only conditions to show a lower Herfindahl score in partner choices and a larger standard deviation in reward estimations than participants in the random-meet condition. Moreover, we predict participants in the yoke-both condition to behave similarly to participants in the self-select condition. Finally, we predict participants in the yoke-choice-only condition to be less biased than participants in the yoke-both condition. In other words, we expect to see graded standard deviations of the estimated rewards, with the self-select condition showing the biggest and the random-meet condition showing the smallest effects. The hypothesis would fail if there are no estimated differences between conditions (see materials and pre-registrations in S2.1-2.2 and S3.2).

Participants

Following a power analysis with pilot data including the two yoked designs (see pilot details in S3.2), we recruited 2005 online workers via Amazon Mechanical Turk's Cloud Research high-quality pool with 500 participants in each of the 4 conditions, in order to obtain small-to-medium effects for the yoked conditions.

Results

As predicted and replicating Experiment 1, participants in the self-select (therefore yoke-both and yoke-choice-only) conditions were more likely to show lower Herfindahl score on partner choices (*b* = -.163, 95% *CI* [-.189, -.138], *p* < .001; μ(random-meet) = .75, δ (random-meet) = 0, *N* (random-meet) = 502, μ(self-select) =.57, δ(self-select) = .24, *N* (self-select) = 502, Cohen's *d* = 1.06) than participants in the random-meet condition, indicating they interacted more selectively. For number of interactions, participants in the random-meet condition interacted 10 times with each group, yet participants in the self-select (therefore yoke-both and yoke-choice-only) condition interacted 20, 10, 8, and 6 times on average with the perceived best, second, third, and worst group respectively (Fig 4a).

Also as predicted and replicating Experiment 1, participants in the self-select condition were more likely to show a larger standard deviation of the estimated rewards than participants in the random-meet condition (*b* = 4.113, 95% *CI* [2.845, 5.382], *p* < .001; μ(random-meet) = 8.82, δ(random-meet) = 8.02, *N* (random-meet) = 502, μ(self-select) = 12.93, δ(self-select) = 10.83, *N* (self-select) = 502, Cohen's *d* = .43), indicating they perceived groups to differ more. As predicted, participants in the yoke-both (*b* = 5.464, 95% *CI* [4.194, 6.734], *p* < .001; μ (yoke-both) = 14.28, δ(yoke-both) = 11.60, *N* (yoke-both) = 500, Cohen's *d* = .55) and yoke-choice-only (*b* = 3.927, 95% *CI* [2.658, 5.196], *p* < .001; μ(yoke-choice-only) = 12.75, δ (yoke-choice-only) = 10.19, *N* (yoke-choice-only) = 501, Cohen's *d* = .43) condition also showed a larger standard deviation than the random-meet condition. In terms of concrete estimations, participants in the random-meet condition estimated the rewards to be on average 90, 84, 77, and 68 (out of 100) points. In contrast, participants in the self-select condition estimated the rewards to be 84, 72, 62, and 52 points (Fig 4b).

Unexpectedly, passive learning exacerbated bias (Fig 4b): Participants in the yoke-both condition estimated the rewards to be on average 89, 75, 65, and 54, which shows a larger standard deviation than the self-select participants (*b* = 1.351, 95% *CI* [.081, 2.621], *p* = .037). Our model would not predict a difference. Active learning research would predict the opposite.

We speculate the under-estimation of the best group in the self-select condition may contribute to this empirical finding.

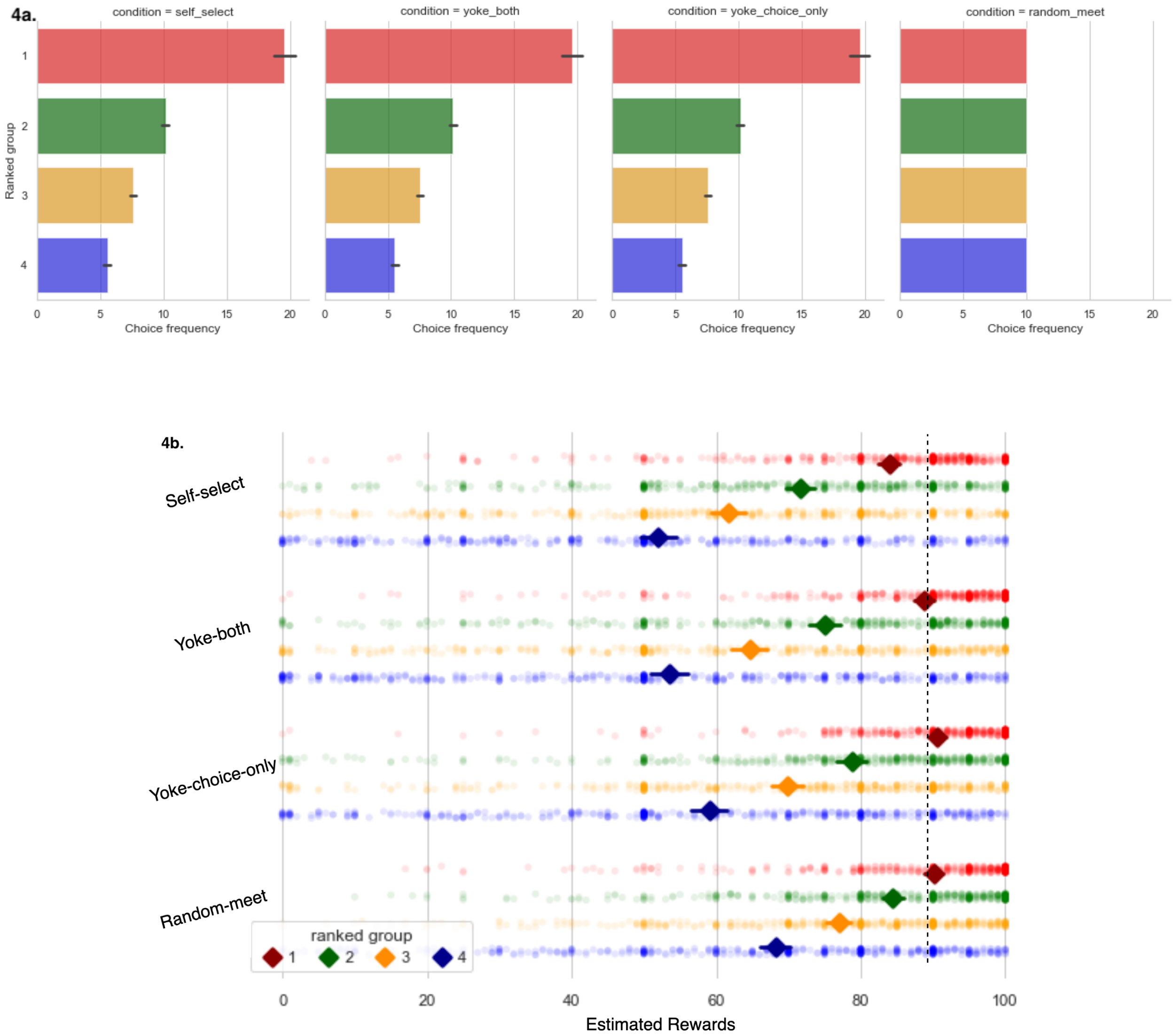


Fig 4. Explore Toma City Experiment 2 testing mechanisms. Plots display empirical results of times of interactions (bar plots with 95% bootstrapped confidence intervals) and estimated rewards in the final round (point estimates with 95% bootstrapped confidence intervals with raw data in the background) for each group. Dashed lines in the bottom subfigure represent true utility. See interpretation in the main text and more analyses in S4.2.

Finally, as expected, reward exposure matters (Fig 4b): Participants in the yoke-choice-only condition estimated the rewards to be on average 91, 79, 70, and 59, which shows a smaller standard deviation than the yoke-both participants (*b* = -1.537, 95% *CI* [-2.807,

-.266], *p* = .018). This is consistent with our model and simulation, such that when the sequence of the rewards is disturbed, estimation will be disturbed as well (see analysis details in S4.2-3 and an exploratory analysis with warmth and competence perceptions in S4.2-4).

Also replicating Experiment 1, an individual-level analysis reveals that the more a participant interacted with one group predominantly, the more likely that participant estimated a larger group difference in rewardingness (Pearson's *r*(2003) = -.397, *p* < .001).

In sum, a higher statistical power replication confirms that participants may have engaged in locally adaptive exploration when exploring a fictional city. As a result, they form inaccurate impressions by perceiving differences between groups, when in reality, there are no group-level differences. The selective samples (rather than the sense of control) and the exposure to rewards (rather than the mere presence of choices) might be key mechanisms that contribute to this phenomenon.

Discussion

Exploring an environment with no differences between groups suffices for inaccurate stereotypes. We offer a plausible, minimal, sufficient means by which stereotypes can develop. Without already demonstrated motivations (from social identity, dominance, or threat), cognitive limitations (from selective attention or prior biases), or information deficits (from majority v. minority representation), adaptive exploration produces a local impression good-enough for present purposes, but ignorant of other possibilities foregone. This process is functional: Adaptive exploration maximizes utility in particular social environments, without prolonging search. But collateral damage to the unexplored groups flows from an otherwise functional approach. This *minimal*, *functional* paradigm plausibly fits the simulation, the human data, and common sense.

This theoretically clean and precisely defined paradigm can guide research on why people form biases. For example, biased impressions may start even before people develop prejudices (e.g., right wing authoritarianism, social dominance orientation), endorse malicious motivations (e.g., active oppression, resource control), experience cognitive limitations (e.g., cognitive miser, attentional shifts), or encounter information deficits (e.g., group size differences). The minimal conditions sufficient for biased impressions to emerge are two: All groups in the environment are equally likely to result in a successful interaction, and decision makers make adaptive choices about those interactions.

Of course, people with prior prejudices, malicious motivations, higher cognitive loads, selective attention, or lower accessibility to one group should be even more, not less, likely to form inaccurate impressions. Research should test how these factors interact with our basic paradigm. For example, illusory correlation (Hamilton & Gifford, 1976; Sherman et al., 2009) has assumed asymmetric group sizes. Our paradigm relaxes that assumption; even if groups are equal in size, illusory correlations can still emerge, as a result of exploratory sampling.

Further insight comes from the observation that both our model and human participants were more likely to form accurate impressions when the underlying rewards were different. This does not necessarily mean that people will never form inaccurate impressions given real group differences. First, the impressions about the least interacted groups were always inaccurate. Second, other mechanisms such as motivational biases and cognitive limitations can still play an important role in those situations. Future work may test more complex combinations.

Another connection is ratio bias, people amplifying proportional trends in large samples (Fiedler et al., 2016). Because the true proportion of rewarding interactions for each group was high in our study, we actually saw the opposite: people reduce proportional trends in small samples. Future work should explore changes in sample size to enrich the current paradigm.

Stereotypes have been considered sometimes as overgeneralizing some kernel of truth (Allport, 1954). Implicit in that definition is the mean difference between groups. However, our work suggests another possibility: Stereotypes do not need to develop based on any kernel of truth; they can be snowballed arbitrariness. In both simulations and experiments, the group with whom the agents interacted the most, and thus estimated the most accurately, is completely arbitrary. It depends on the initial interactions, which by construction are random. However, this limits our analysis to individual-level stereotypes, which differ from collective stereotypes, where society has consensus. Future work can include contextual bandits to examine whether people learn mappings between features that make them more likely to generalize certain stereotypes to certain groups (Schulz et al., 2018). Transmission of information between agents and across generations could be another mechanism to get collective stereotypes (Martin et al., 2014). Stereotypes about social groups are not unidimensional. Future work can test how this paradigm applies to the emergence of complex stereotype contents such as warmth/ communion and competence/agency (Abele et al., 2020).

Understanding the origin of inaccurate stereotypes as one consequence of locally adaptive exploration provides new theoretical and pragmatic insights. Theoretically, our paradigm pays homage to Tajfel's (1971) minimal-group paradigm, identifying minimal conditions for stereotypes to emerge. We predict stereotypes can result from an adaptive solution to a particularly challenging environment (Anderson, 1991). The environment is challenging exactly because social groups can be equally rewarding, so people have multiple ways to maximize long-term rewards from interactions. This goal is at odds with another goal, forming accurate impressions. In seeking to maximize rewards, people are driven to strategies that only explore locally -- focusing on options already considered -- and it is this adaptive exploration that leads to inaccurate impressions. This perspective aims not to justify stereotypes, but on the contrary, to demonstrate an even harder, seemingly innocent, but almost unavoidable route for biases to emerge. This perspective also encourages researchers to think about how to better modify the social environment and exploration strategies to reduce inaccurate stereotypes.

Pragmatically, our paradigm predicts that even if one could address the subpopulation who have malicious intentions or cognitive limitations or if the environment equips people with equal information about all groups, social psychologists should not expect inaccurate stereotypes will naturally disappear. To reduce inaccurate stereotypes, we should first intervene at the environmental level. For example participants in the random-meet condition were exposed to diverse representations of Toma people, so they formed less biased impressions, as compared to those in the self-select condition with predominantly one group. Correlational studies show consistent evidence that representational diversity reduces perceived differences among groups (Bai, Ramos, & Fiske, 2020), but more natural experiments need to test this hypothesis.

Our analyses also predict a role for strategies that encourage exploration of the environment. Even when the underlying population in a place is diverse, if people just explore based on their past experiences, they will be more likely to behave as our self-select participants do, without noticing other groups can be equally good. Hence, interventions that motivate exploration can combat that all-too-human nature. From this perspective, the intergroup contact hypothesis may fulfill this cognitive mechanism by encouraging people to explore outgroup members (Allport, 1954). Again, more direct experiments and measures need to test this hypothesis.

Drawing on theoretical analysis, simulations, and empirical data, we propose inaccurate stereotypes can result from locally adaptive exploration. Far from a rebuttal of prior explanations, we provide a complementary perspective, a process that is general and deep, independent of group motives, cognitive limits, or information constraints. Perhaps one origin of stereotypes is much simpler than we have thought; even minimal process assumptions can recreate it. However, this simplicity is also troubling: If stereotypes can result from each person pursuing their own self-interest, we may need to work harder to create environments where problematic stereotypes do not develop.

References:

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290.

Agrawal, S., & Goyal, N. (2012, June). Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference on learning theory* (pp. 39-1). JMLR Workshop and Conference Proceedings.

Allport, G. W. (1954). *The nature of prejudice.* Addison-Wesley.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429.

Bai, X., Ramos, M. R., & Fiske, S. T. (2020). As diversity increases, people paradoxically perceive social groups as more similar. *Proceedings of the National Academy of Sciences*, *117*(23), 12741-12749.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate?. *Journal of social issues*, *55*(3), 429-444.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31,* 21–32.

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological review*, *112*(4), 951.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523-538.

Fetchenhauer, D., & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science*, *21*(2), 189-193.

Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hutter, M. (2016). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition, 44(1), 143-161.*

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*(6), 889–906.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Fiske, S. T., & Durante, F. (2016). Stereotype content across cultures: Variations on a few themes. In M. J. Gelfand, C.-Y.Chiu, & Y.-Y. Hong (Eds.), *Handbook of Advances in Culture and Psychology* (Vol. 6, pp. 209-258). New York: Oxford University Press.

Fiske, S. T., & Taylor, S. E. (1984). Social cognition. Reading, Mass: Addison-Wesley.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition, 173,* 34-42.

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, 103948.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392-407.

Harris, C., Fiedler, K., Marien, H., & Custers, R. (2020). Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments. *Journal of Experimental Psychology: General*.

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual review of psychology*, *47*(1), 237-271.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system‐justification and the production of false consciousness. *British journal of social psychology*, *33*(1), 1-27.

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the “naivety” assumption in sampling explanations of judgment biases. *Psychological review*, *118*(2), 379.

Lockwood, P. L., Apps, M. A., & Chang, S. W. (2020). Is there a ‘social’ brain? Implementations and algorithms. *Trends in Cognitive Sciences*.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology*, *51*(1), 93-120.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

Martin, D., Hutchison, J., Slessor, G., Urquhart, J., Cunningham, S. J., & Smith, K. (2014). The spontaneous formation of stereotypes via cumulative cultural evolution. *Psychological Science*, *25*(9), 1777-1786.

Prager, J., Krueger, J. I., & Fiedler, K. (2018). Towards a deeper understanding of impression formation—New insights gained from a cognitive-ecological perspective. *Journal of personality and social psychology*, *115*(3), 379.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2018). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(6), 927–943.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*(1), 131–142.

Stangor, C., & Schaller, M. Ch 1 in Macrae, C. N., Stangor, C., & Hewstone, M. (Eds.). (1996). *Stereotypes and stereotyping*. Guilford Press.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. Journal of Personality and Social Psychology, 96(2), 305–323.

Sherman, S. J., Sherman, J. W., Percy, E. J., & Soderberg, C. K. (2013). *Stereotype development and formation*. Oxford University Press.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression.* Cambridge University Press.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178.

Tajfel, H., & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In S. Worchel, & W. G. Austin (Eds.), The Social Psychology of Intergroup Relations (pp. 33-47). Monterey, CA: Brooks/Cole.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3/4), 285-294.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, *9*(2p2), 1.

Supplemental Online Materials
for
Globally Inaccurate Stereotypes Can Result from Locally Adaptive Exploration

Authors

Note: Please use this document in parallel with the codebook published on OSF (link) and the main text. This file is designed to be used as a reference for readers seeking information on specific topics.

Table of Contents:

S1.1. The primary model: Multi-armed bandit with Thompson sampling

The multi-armed bandit problem has been studied for decades in statistics, operation research, electrical engineering and computer science (see review in Sutton & Barto, 2018). The name comes from a motivating story that a gambler enters a casino and sits down at a slot machine with multiple arms that can be pulled. When pulled, an arm produces a random payout drawn independently of the past. Because the distribution of payouts corresponding to each arm is not known, the gambler can learn it only by experimenting. As the gambler learns about the arms' payouts, he faces a dilemma: He can either exploit arms that yielded high payouts in the past or he can explore alternative arms that may earn higher payouts in the future. What is the best strategy for pulling arms that balances this tradeoff and maximizes the cumulative payouts?

Thompson sampling (see review in Russo et al., 2017), also known as posterior sampling or probability matching, was first introduced for allocating experimental effort in two-armed bandit problems arising in clinical trials. The algorithm was largely ignored until recently, with a surge of interest in industry and academia given its strong empirical performance. Here we briefly introduce Thompson sampling for the Beta-Bernoulli multi-armed bandit.

There are *K* arms. When played, an arm *k* produces a reward of 1 with probability $\theta_k$and a reward of 0 with probability $1 - \theta_k$. Each $\theta_k$can be seen as an arm's mean reward. The mean rewards $\theta = (\theta_1,..., \theta_k)$ are fixed but unknown. In the first period, an action $x_1$is applied, and a reward $r_1 \in \{0, 1\}$is generated with success probability $P(r_1 = 1 \mid x_1, \theta_1) = \theta_{x1}$. After observing r_1, the player applies another action x_2, observes a reward r_2, and this process continues. Let the player begin with an independent prior belief over each $\theta_k$expressed as a beta distribution parameterized with $\alpha = (\alpha_1,..., \alpha_k)$ and $\beta = (\beta_1,..., \beta_k)$. For each arm *k*, the prior probability density function of $\theta_k$is

$$p(\theta_k) = \Gamma(\alpha_k + \beta_k) / \Gamma(\alpha_k)\Gamma(\beta_k)\, \theta_k^{\alpha_k - 1} (1 - \theta_k)^{\beta_k - 1},$$

where Γdenotes the gamma function. As the player gathers more observations, this distribution is updated according to Bayes' rule. The beta distribution is convenient to work with because of its conjugacy properties. Each arm's posterior distribution is also a beta distribution, updated according to the rule

$$(\alpha_k, \beta_k) \leftarrow (\alpha_k, \beta_k), \text{if } x_t \neq k$$

$$(\alpha_k, \beta_k) \leftarrow (\alpha_k, \beta_k) + (r_t, 1 - r_t), \text{if } x_t = k$$

Note that only the parameters of a selected arm are updated. A beta distribution with parameters (α_k, β_k) has mean $\alpha_k / (\alpha_k + \beta_k)$, and the distribution becomes more concentrated around this mean as $\alpha_k + \beta_k$grows.

Procedurally, in each time period *t*, the algorithm generates an estimate $\widehat{\theta_k} = \alpha_k / (\alpha_k + \beta_k)$, equal to its current expectation of the success probability θ_k. The success probability estimate $\widehat{\theta_k}$ is randomly sampled from the posterior distribution. Thus, $\widehat{\theta_k}$ represents a statistically plausible success probability rather than a statistically plausible observation. The arm $x_k$with the largest estimate $\widehat{\theta_k}$ is then applied, and the player observes a new reward r_{t+1}.

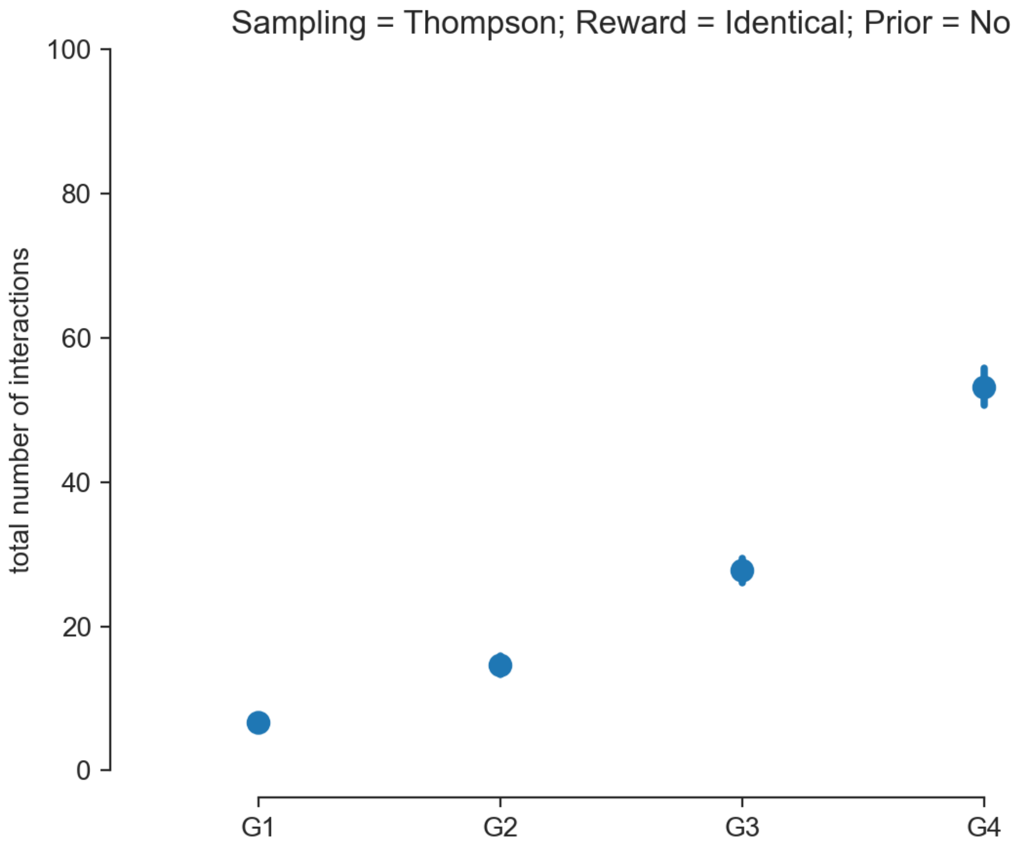
S1.1-1a. Pseudocode for Thompson sampling for Bernoulli bandits.

```
Algorithm: Thompson Sampling Bernoulli Bandits

for each arm $k = 1, 2, \ldots, K$, set $\alpha_k = 1,\ \beta_k = 1$.

for each round t = 1, 2, …, T, do:
    for each arm k = 1, …, K, sample $\theta_i(t)$from the Beta distribution $(\alpha_k, \beta_k)$.
    play arm $k(t)$ : = argmax $\theta_k$, and observe reward $r_t$.
    if $r_t$=1, then $\alpha_{k(t)} = \alpha_{k(t)} + 1$, else $\beta_{k(t)} = \beta_{k(t)} + 1$.
end
```

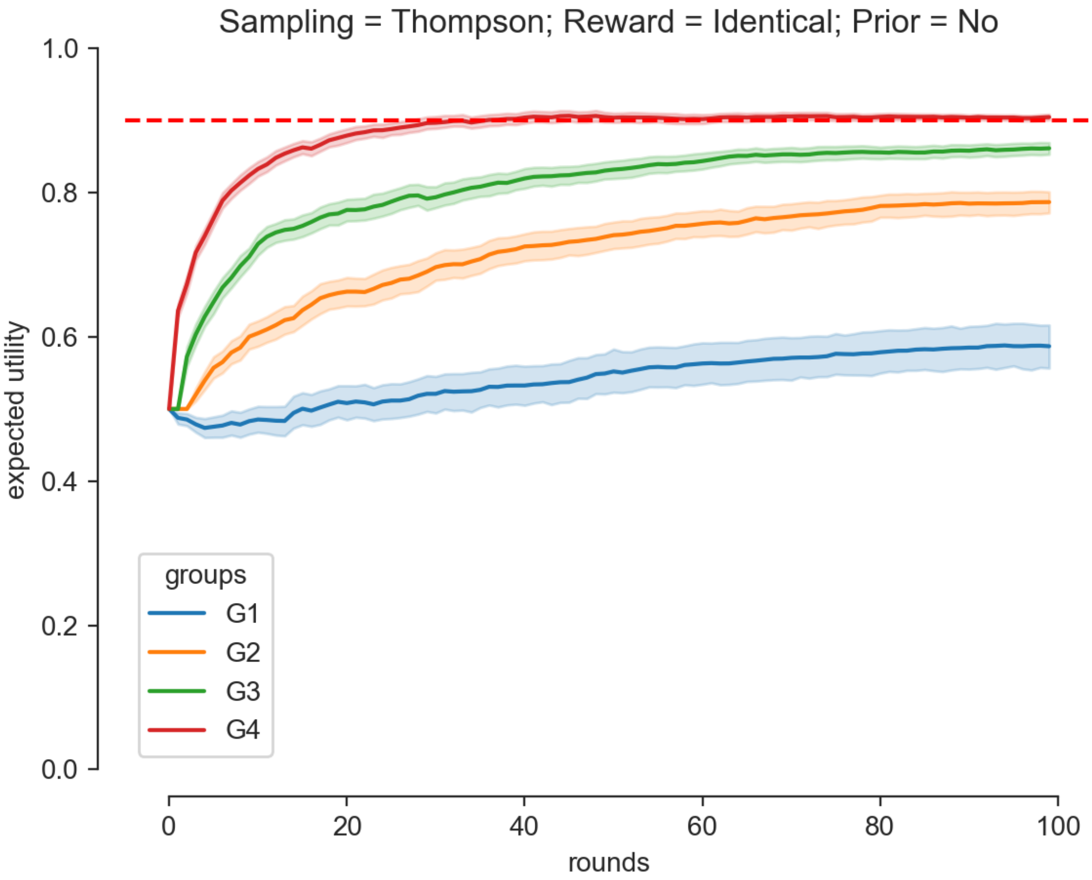
S1.2. Simulation results: Longer timescales

The main text reported results from 50 simulations with 4 groups over a 40-round game. This is consistent with Experiment 1. Here we supplement this with 100 simulations, each with 4 groups over a 100-round game for generality. We also provide more notation to interpret the results. S1.2-1a and S1.2-1b are our main hypothesis/baseline conditions. Please compare each of the following Figures (S1.2-2a, -2b, -3a, -3b, -4a, -4b) against this baseline.

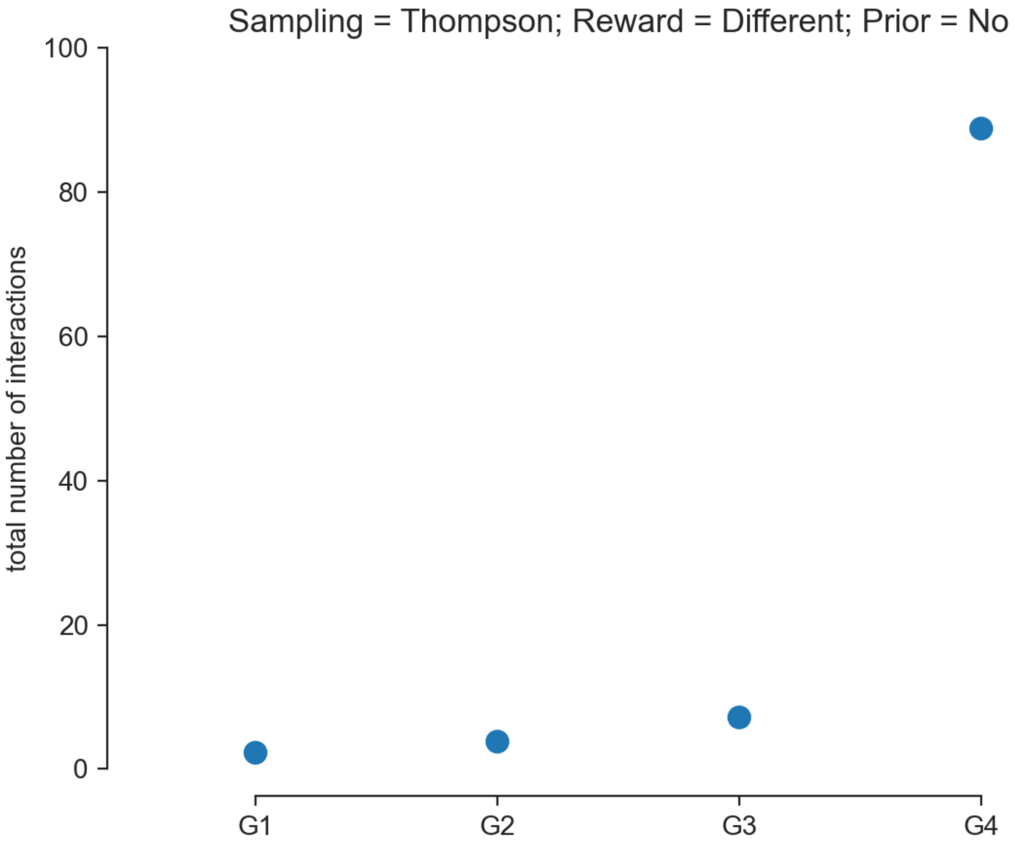
S1.2-1a. Total number of interactions indicates the Thompson sampling algorithm, under the identical and high reward environment with no prior biases, interacts more with one group (i.e., G4) but less with the other groups (i.e., G1, G2, G3). Each blue dot indicates the mean times of interaction across simulations with 95% bootstrapped confidence intervals representing uncertainty.



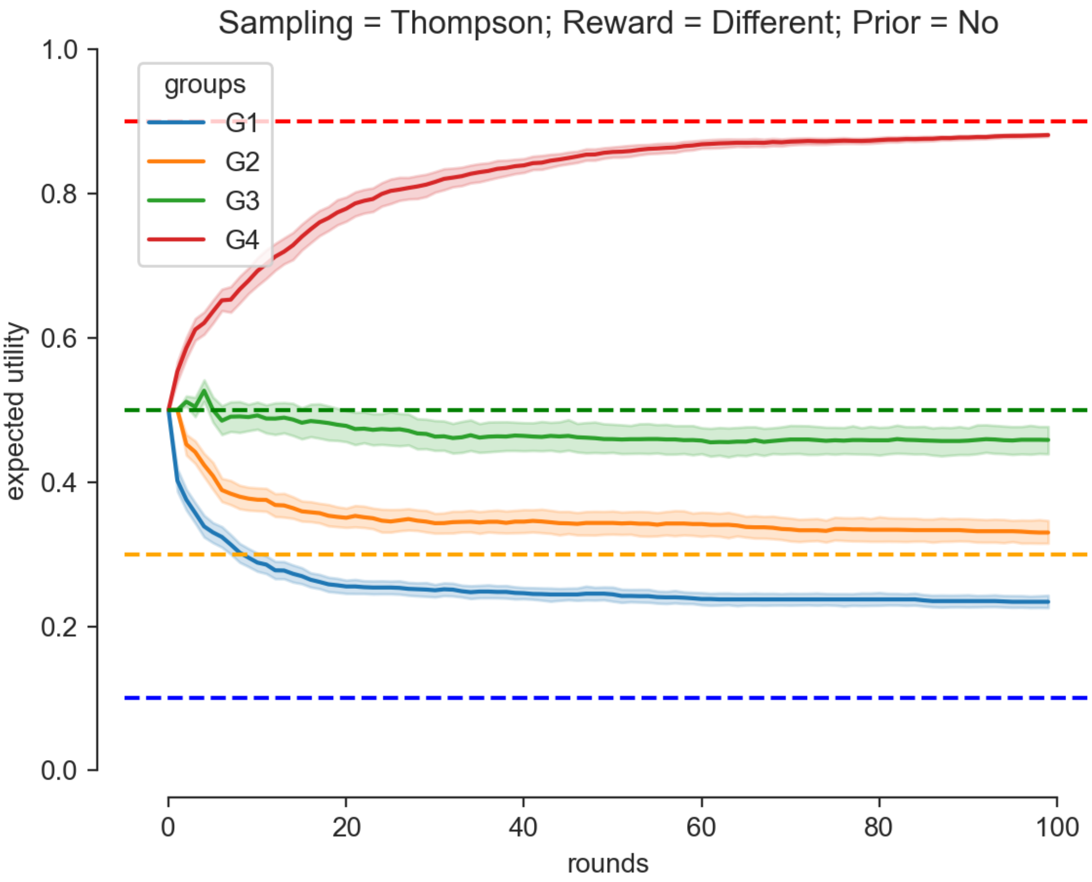
S1.2-1b. Expected utility for each group indicates the Thompson sampling algorithm, under the identical and high reward environment with no prior biases, estimates accurately for the group it interacts the most (i.e., G4) but it estimates much less accurately for the other groups it interacts less (i.e., G1, G2, G3). The red dotted line indicates the ground truth utility for each group (.9), but we see only estimations about G4 approaches the ground truth. Each line indicates the expected utility after each interaction over the course of 100 interactions with 95% bootstrapped confidence intervals representing uncertainty.



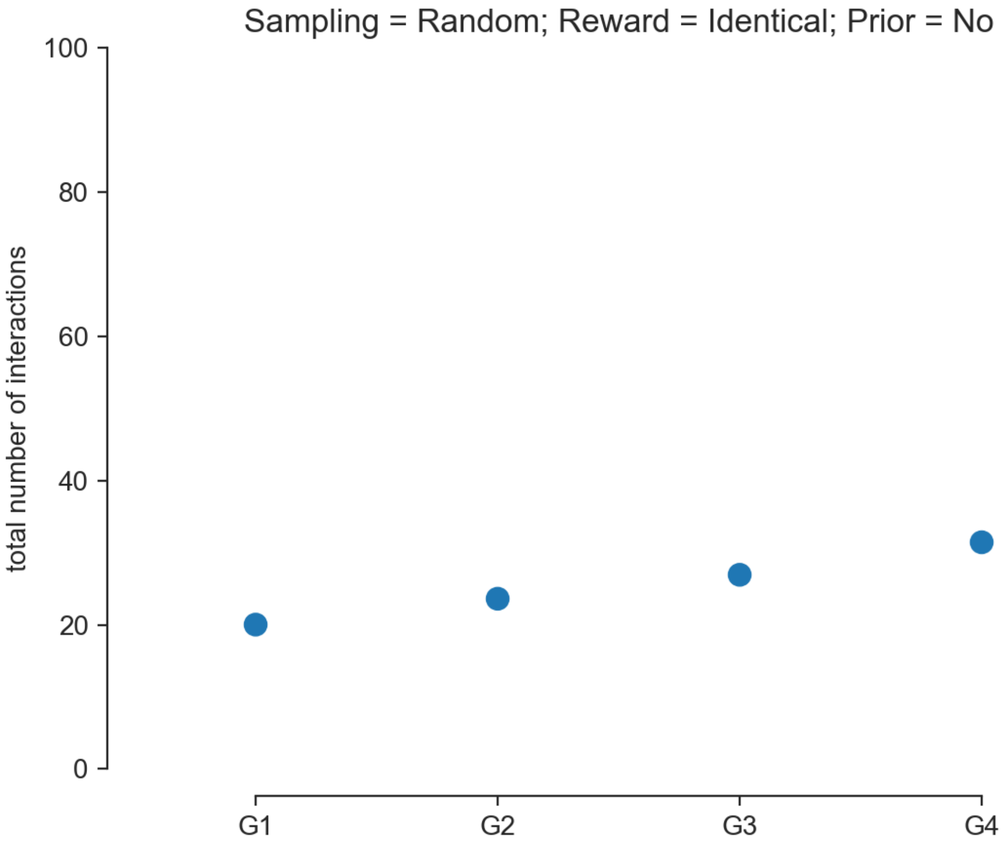
S1.2-2a. Total number of interactions indicates the Thompson sampling algorithm, under the different reward environment with no prior biases, behaves as expected. It interacts more with the group that has the highest expected utility (i.e., G4) and significantly less with the other groups (i.e., G1, G2, G3). As if it finds the "best" arm and sticks to it in order to get higher payoffs.



S1.2-2b. Expected utility for each group indicates the Thompson sampling algorithm, under the different reward environment with no prior biases (i.e., a standard bandit problem), estimates accurately for the group it interacts the most (i.e., G4), and relatively accurately for the second and third groups (i.e., G3 and G2), but it estimates less accurately for the group it interacts least (i.e., G1) due to insufficient sampling.

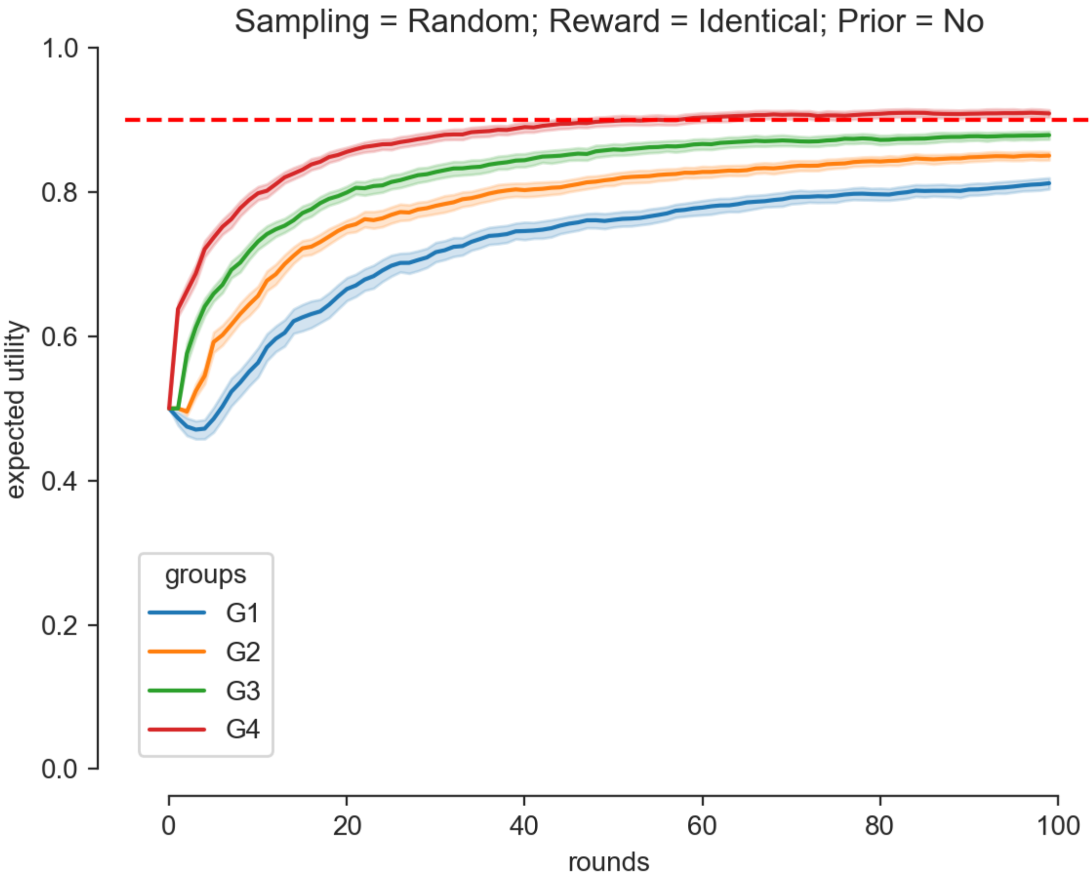


S1.2-3a. Total number of interactions indicates the Random sampling algorithm, under the identical and high reward environment with no prior biases, behaves randomly. It interacts roughly equally with each of the four groups.

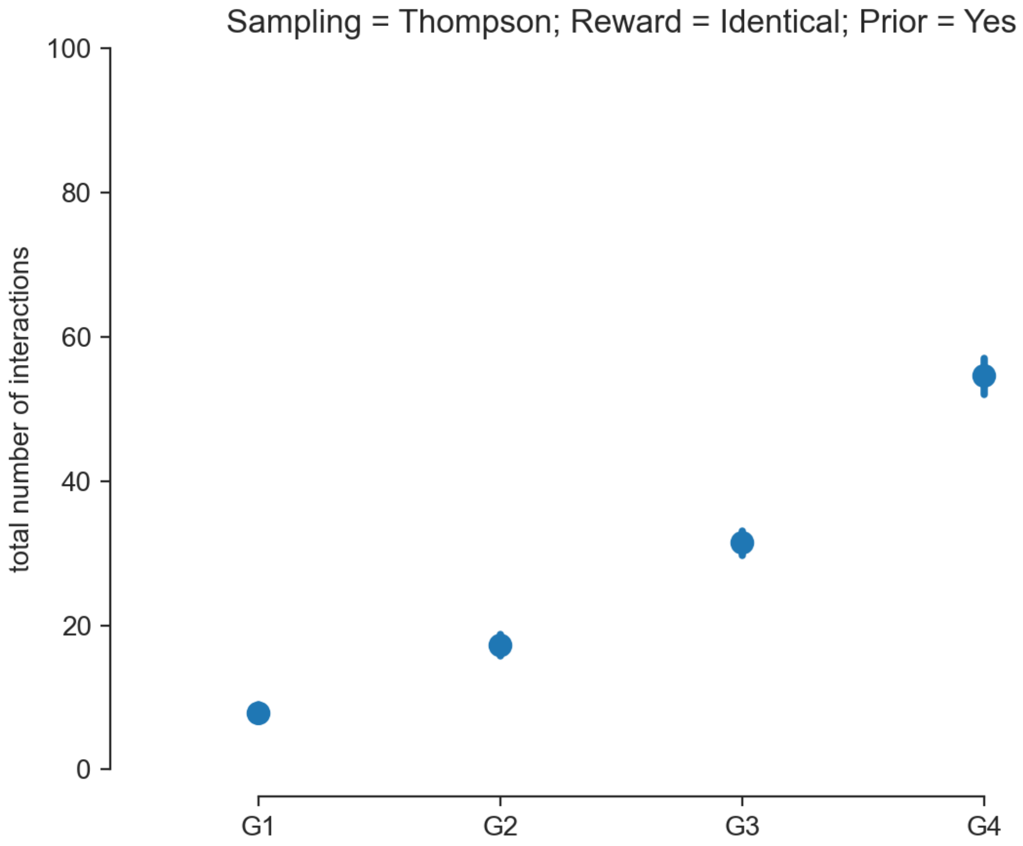


S1.2-3b. Expected utility for each group indicates the Random sampling algorithm, under the identical and high reward environment with no prior biases, estimates accurately for all four groups. It gives more accurate estimates for G1, G2, and G3, as compared to Thompson sampling.

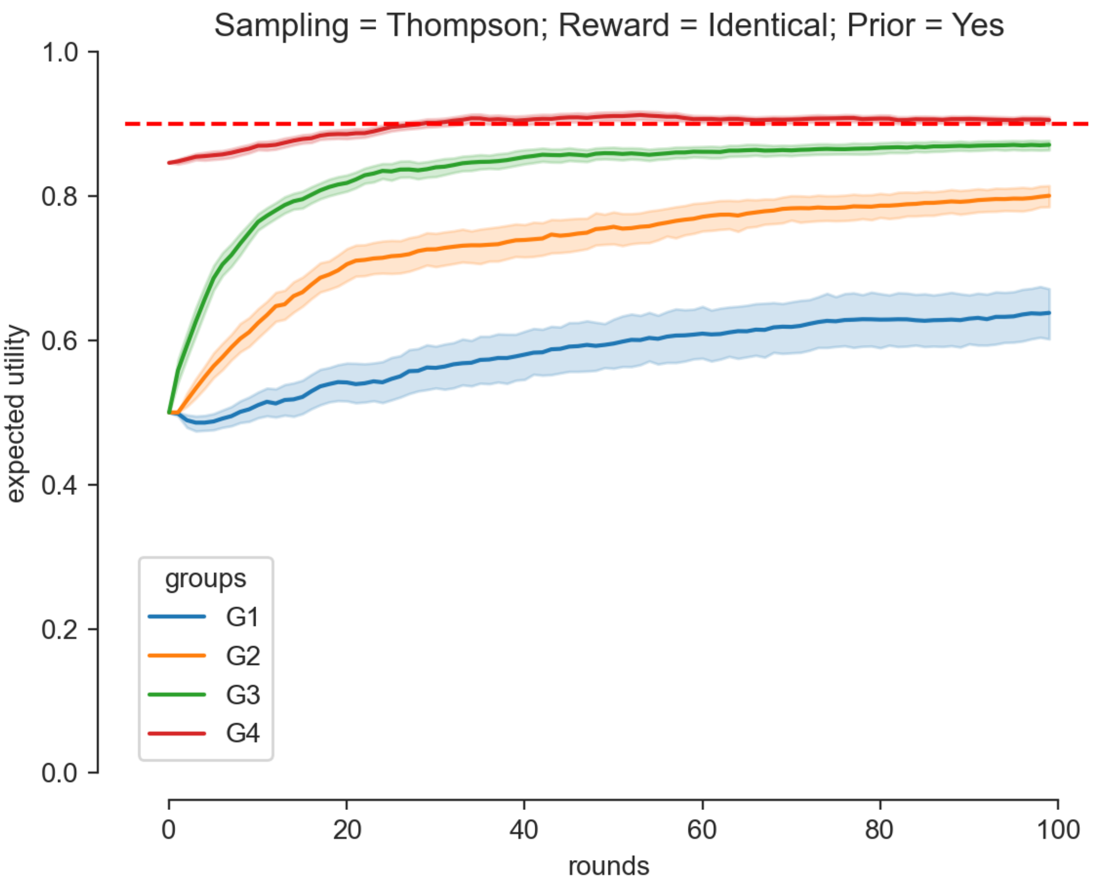
Note a small difference in estimated utility at round 40 even with random sampling strategy. The reason for such difference comes from an artifact of order statistics. For each simulation, we rank-ordered the responses from the most interacted group and the highest expected utility to the least (pre-registered). One consequence of this ordering is that we see maximum numbers always go together, and vice versa minimum numbers. For this reason, when we interpret differences in order statistics we need to do so between groups (random vs. Thompson).



S1.2-4a. Total number of interactions indicates under the identical and high reward environment, small prior biases ($\alpha_k = 10,\ \beta_k = 1$) makes the Thompson sampling algorithm to interact more with that positively-biased group (i.e., G4).



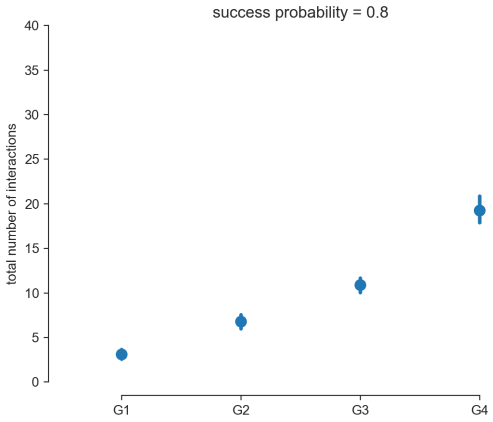
S1.2-4b. Expected utility for each group indicates the Thompson sampling algorithm is more accurate when estimating the positively-biased group (i.e., G4), but it has little influence on the other three groups. It still estimates less accurately for the other three groups.

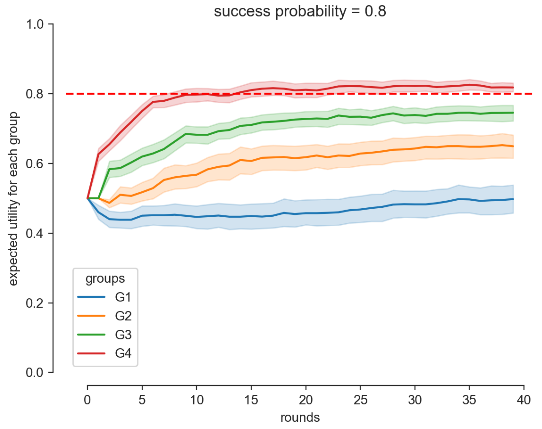


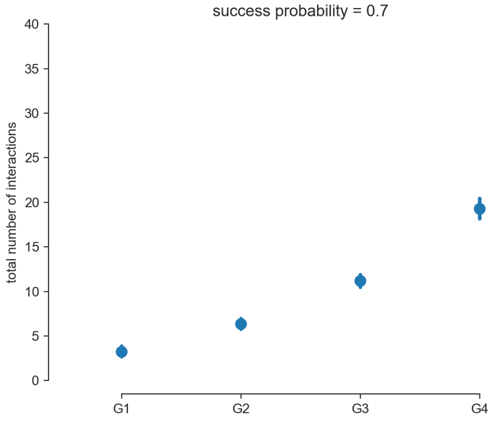
S1.3. Simulation results: Varying the ground truth success probabilities

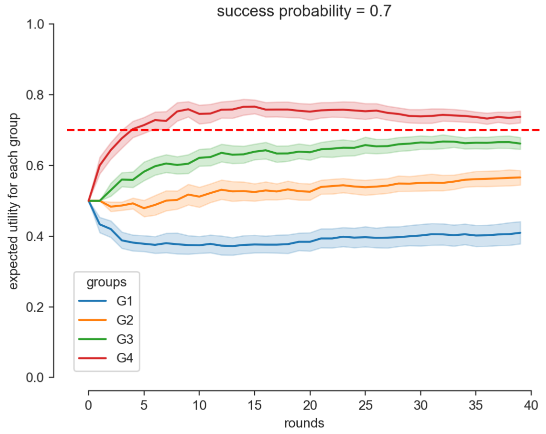
Here, we show simulation results for different success probabilities. In the main text, we provided an identical and high reward environment, $\theta_k = 0.9$. For exploratory purposes, we simulate what happens when $\theta_k$is lower, specifically $\theta_k = 0.8,\ 0.7,\ 0.6,\ 0.5,\ 0.4,\ 0.3,\ 0.2,\ 0.1$. Again, we plot times of interaction and expected utility for each of the four groups.

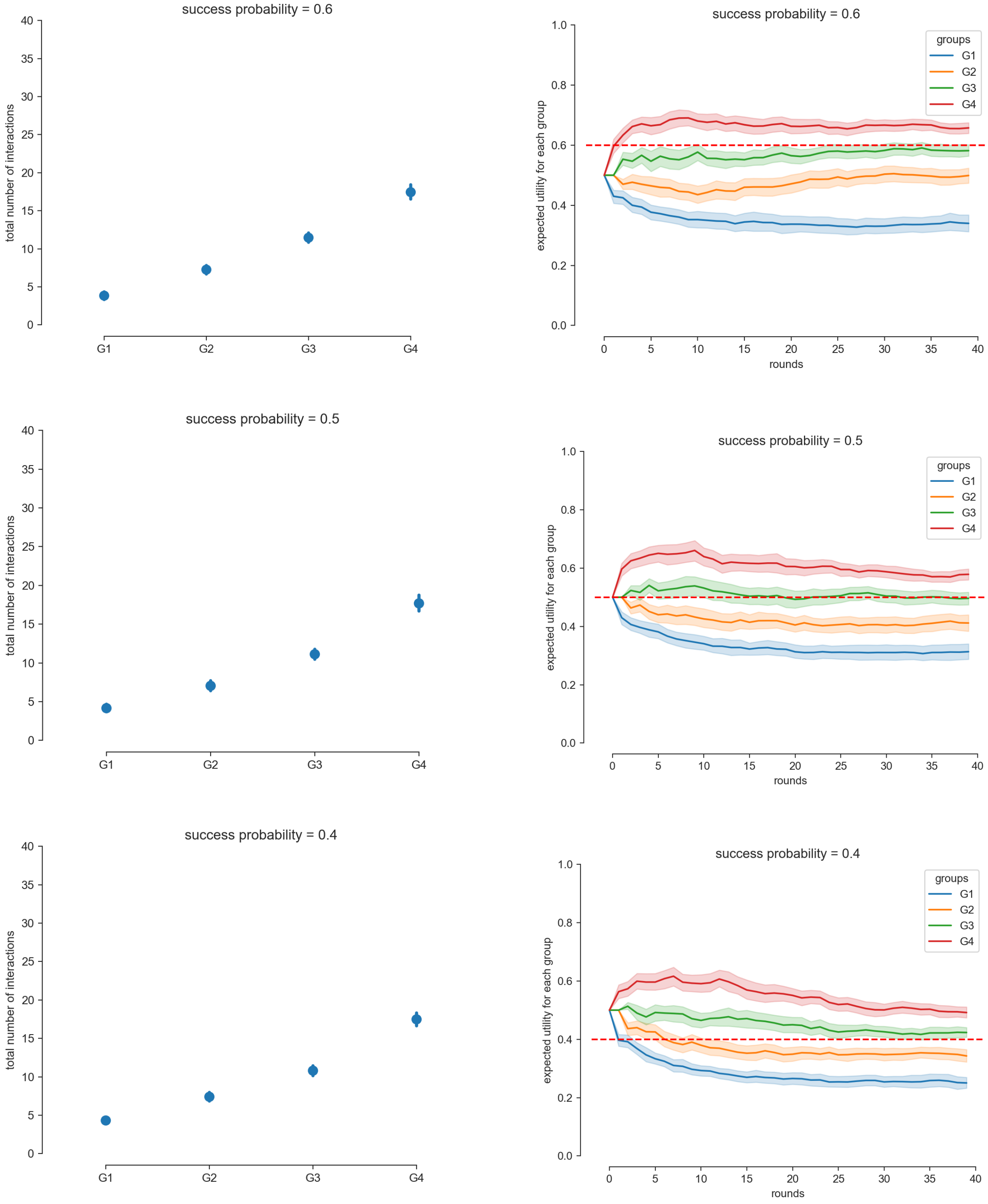
Comparing across success probabilities, we observe that the higher the success probability the more biased interactions and the larger estimated rewards (e.g., $\theta_k = 0.9$ or 0.8 vs. $\theta_k = 0.1$or 0.2). This is consistent with a recent paper (Harris et al., 2020) which suggests biased priors could be consolidated only in reward-rich but not reward-impoverished environments. However, here we found even without prior biases, we still see more biased interactions and biased estimations to emerge.

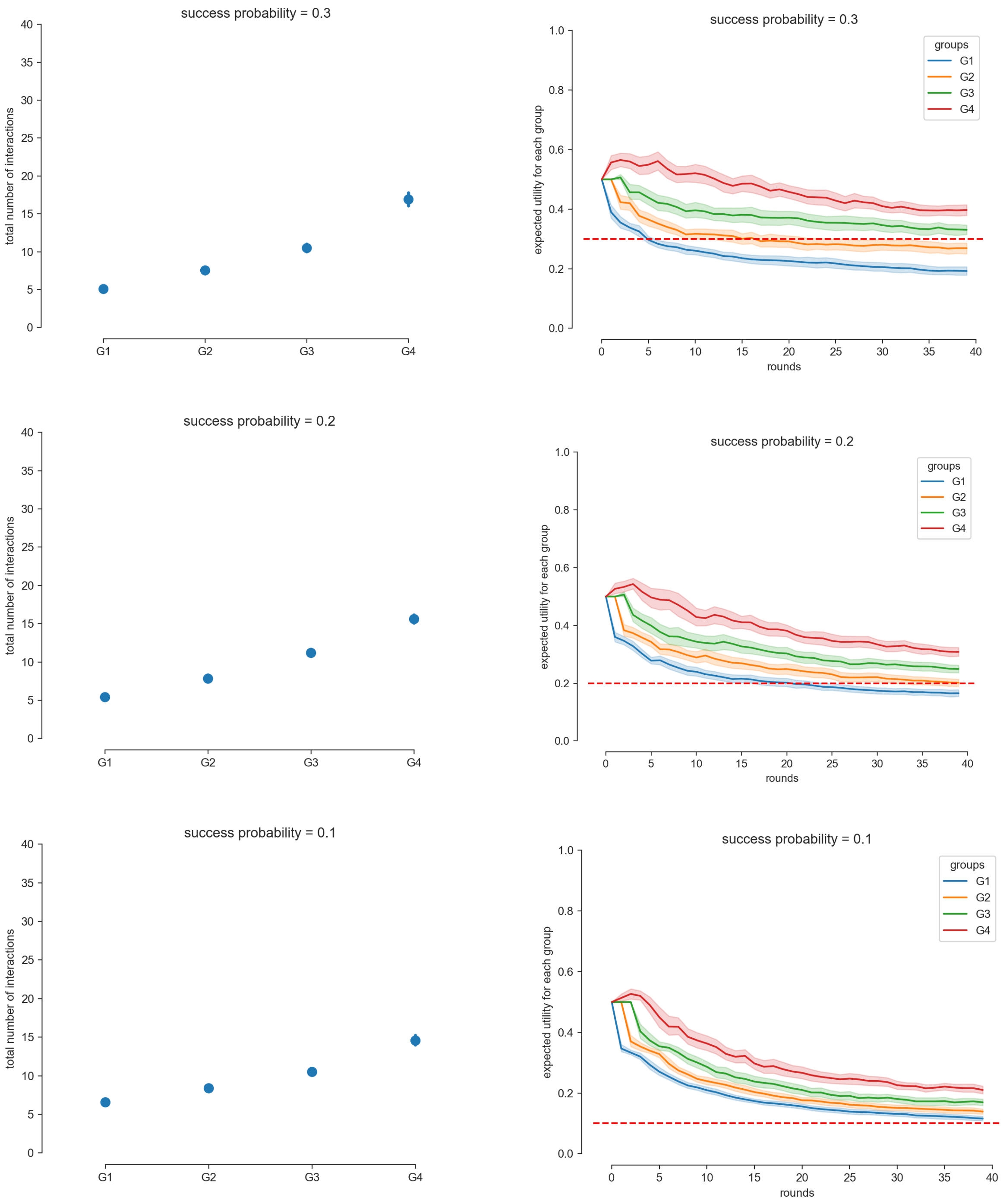








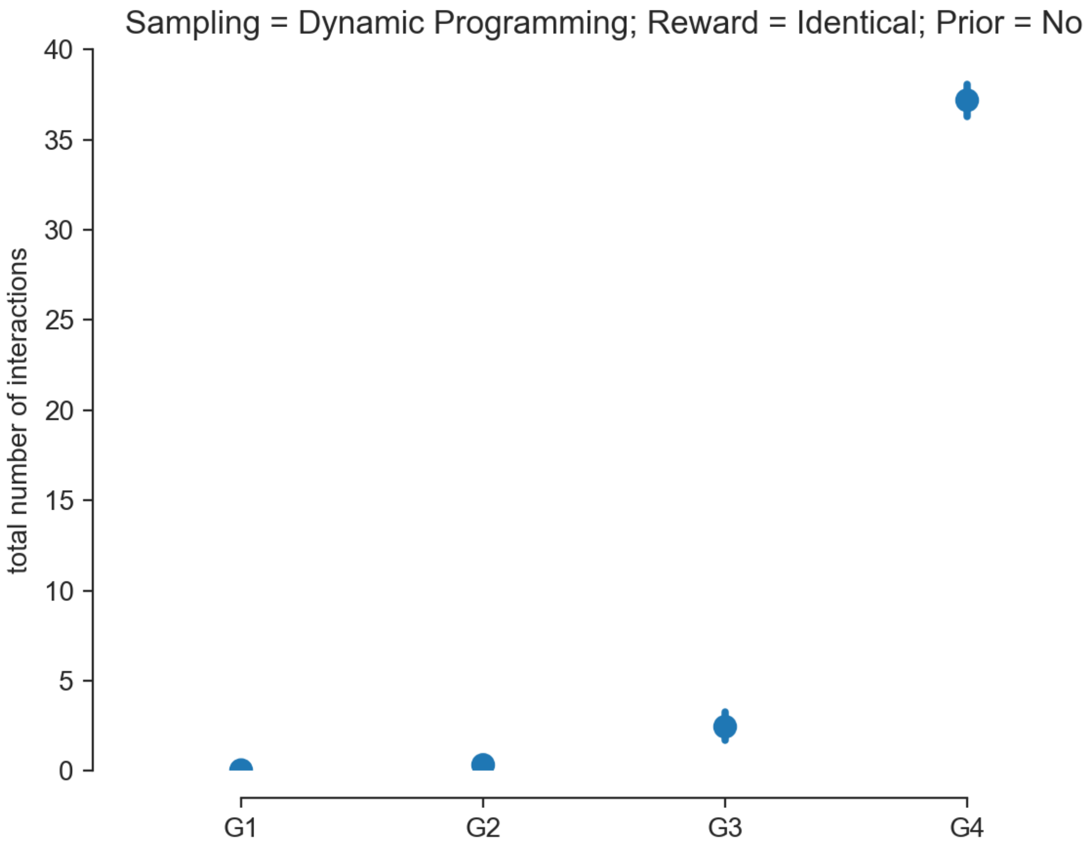

success probability = 0.6
total number of interactions
0
5
10
15
20
25
30
35
40
G1
G2
G3
G4
success probability = 0.6
expected utility for each group
0.0
0.2
0.4
0.6
0.8
1.0
rounds
0
5
10
15
20
25
30
35
40
groups
G1
G2
G3
G4
success probability = 0.5
total number of interactions
G1
G2
G3
G4
success probability = 0.5
expected utility for each group
rounds
groups
G1
G2
G3
G4
success probability = 0.4
total number of interactions
G1
G2
G3
G4
success probability = 0.4
expected utility for each group
rounds
groups
G1
G2
G3
G4


success probability = 0.3
total number of interactions
0
5
10
15
20
25
30
35
40
G1
G2
G3
G4
success probability = 0.3
expected utility for each group
0.0
0.2
0.4
0.6
0.8
1.0
rounds
0
5
10
15
20
25
30
35
40
groups
G1
G2
G3
G4
success probability = 0.2
total number of interactions
success probability = 0.2
expected utility for each group
rounds
groups
success probability = 0.1
total number of interactions
success probability = 0.1
expected utility for each group
rounds
groups

S1.4. An alternative model: Multi-armed bandit with dynamic programming

We focused on Thompson sampling, which minimizes the cumulative errors of an agent interacting with a multi-armed bandit for potentially infinitely many rounds. For a finite horizon (ie. a fixed number of rounds) multi-armed bandit, Thompson sampling is a heuristic strategy and the optimal solution can be computed by dynamic programming (Bellman, 1957). This algorithm is based on the principle that if the optimal policy from a certain stage onward is known, then it is relatively easy to extend this policy so as to give an optimal policy starting one stage earlier. Repetition of this procedure derives the optimal policy. All simulations are run in the environment with success probability identical and high as in the main text.

S1.4-1a. Simulation results for dynamic programming. Compared to our baseline Thompson sampling algorithm (40 rounds of game with 4 arms, identail and high success probabilities at .9 without prior bias), the optimal solution derived from dynamic programming, as shown below in Figure S1.4-1a, is more extreme: it is much more likely to interact with only one group and much less likely to explore other groups.



S2.1. Human experiment infrastructure: materials and logics

The human participant experiment recreates a situation where people travel to a new place and meet some new people. To avoid any non-experimentally induced prior beliefs, we used a fictional city called Toma City and four fictional social groups: Tufa, Aima, Reku, and Weki. Human participants are invited to play games with people in Toma City.

Below is a verbatim description of study materials. **Boldfaced** texts are experiment phases and treatments. Yoked designs are engineered in the following manner: We first collected responses from the self-select condition. We then stored all responses (i.e., choices) from the self-select condition on a server. We next prepared study materials for the yoked conditions (e.g, you meet instead of you choose), and for each new incoming participant, we fetched stored responses from the server as yoke-stimuli to show to that participant. Hence, participants arrived in sequential order, meaning the next participant only joined the study after the prior participant completed the study. This makes sure the server writes in and writes out the correct response (i.e., not showing one identical response to two or more participants). See JavaScript in our online repository.

Admittedly, such a procedure yields one potential confound to random assignment -- the time of arrival. Participants in Experiment 2 self-select and random-meet conditions arrived (clicked the study link) approximately eight hours earlier than yoke-both and yoke-choice-only conditions. If there are systematic differences before and after the eight-hour window, it would create a violation of random assignments. Due to platform limitations, we could not establish a fully randomized experiment. However, we made sure to exclude double-participation to minimize potential confounds, as well as double-checked participant demographics which yielded similar compositions in terms of gender, race, and age.

- Self-select: 55% female, 82% White, and 73% 20-40yrs;
- Random-meet: 53% female, 79% White, and 70% 20-40yrs;
- Yoke-both: 49% female, 83% White, and 83% 20-40yrs;
- Yoke-choice-only: 56% female, 82% White, and 81% 20-40yrs.

Cover:
Generate underlying reward distributions: 0.1, 0.3, 0.5, 0.9 vs. 0.9, 0.9, 0.9, 0.9
They read:
In this study, you will play a game with made-up people from a made-up city. Toma City has around 100,000 residents. These residents come from 4 ancestral villages: Tufa, Aima, Reku, Weki. Let's first get to know the people in Toma City! Each village group has its own symbol. Tufa is Fire. Aima is Gold. Reku is Rock. Weki is Water. Let's now play a game with them!

Prior biases:
Show information v. no show
They read:
According to some locals in Toma City: Tufas don't have a lot, but they are willing to help others. Aimas are rich and a little selfish. Rekus are wealthy, and very generous to their neighbors. Wekis are very poor; they are not interested in other people or the community.

Game instruction:
Instruction for self-select v. random-meet (i.e, yoke-both and yoke-choice-only)
They read:
Collaborate with Toma; Earn more points. To begin with, you will have 10 points. You can increase your points by starting a small business in the city. You can **select [meet]** some people to help you. Who is going to help you? Some people are able to increase your points and are willing to share the profits with you. But some will neither increase nor share the points. Who will you collaborate with? You will play for 40 rounds. Each round you get to **choose [meet]** one person, with a new set of people to **choose [meet]** each time. If the person you chose helped you grow your business, you earn 1 point. If not, you earn 0 points. You will see your points after each decision. The more points you earn, the more bonus you get. (1 point = 1 cent). Ready to play? Let's meet some new Tufas, Aimas, Rekus, and Wekis!

Game test trials (loop):
They read:
Round **1** of 40.
Find the best person; Earn more points! A new Tufa, Aima, Reku, and Weki just joined you.

Self-select: You choose a [ ]. Tell us, briefly, why did you choose that person: [ ].
(Note: allow click in the self-select condition only, but view-only in other conditions)
Random-meet (i.e., Yoke-both/Yoke-choice-only): You meet a [ ]. Tell us, briefly, what do you think about this choice: [ ].
(Note: choices for random-meet are from initially defined reward generators, on average 10 times each group but in randomized order, whereas choices for the yoke conditions are fetched from the server).

Nice. You collaborated with a [ ]. You earned [ ] point.
(Note: rewards for the self-select, random-meet, and yoke-choice-only conditions are from initially defined reward generators, whereas rewards for the yoke-both condition are fetched from the server).

Estimations:
They read:
Thank you for playing with the Tomas. We hope you enjoyed this game! Now if you have to guess: for each group, how many times out of 100 do you think working with a person from that group would result in you earning 1 point? On a slider from 0 (always lose) to 100 (always win) with intervals of 1, for Tufa, Aima, Reku, and Weki, respectively.

They read:
If your friends are going to visit Toma City. What advice would you give? They write open-ended responses forTufa, Aima, Reku, and Weki, respectively.

They read:
Finally, please rate each group on the following traits: Tufas are warm. Tufas are competent. On a Likert scale from 1 (not at all) to 5 (extremely). Likewise for Aima, Reku, and Weki.

S2.2. Demo videos: Show live experiment from each condition

Here we show video clips (a shorter version with 5 test trials) for the self-select and random-meet conditions. Yoked designs have an experience equivalent to the random-meet condition, so we omitted them here.

- Self-select condition see here.
- Random-meet see here.
- Dependent measures in all conditions (after the game) see here.

S3.1. Human experiment pilot and pre-registrations: Pilot 1

With the same experiment material, we conducted a pilot study with *N* = 30 per condition and 8 conditions in total. This pilot data suggested a sample size for partner choice to be 26 per cell giving an effect size of *d* = .8 at alpha = .05 and beta = .80 level.

We found the average value of partner choice Herfindahl score for the self-select condition was .52 (sigma = .01, effective *N* = 33) and for the random-meet condition was .75 (sigma = .00, effective *N* = 27). These empirical values give us an estimated effect size of *d* = 1.0. We took a conservative number of the largest effect size of .80. The same pilot data suggests a sample size for reward estimation to be 30 per cell gives an effect size of *d* = .8 at alpha = .05 and beta = .80 level.

We found the average value of estimated reward standard deviation for the self-select condition was 7.87 (sigma = 7.66, *N* = 33) and for the random-meet condition was 15.39 (sigma = 12.33, *N* = 27). These empirical values give us an estimated effect size of *d* = .75. Therefore, we decided a-priori to collect 30 participants per condition for Experiment 1.

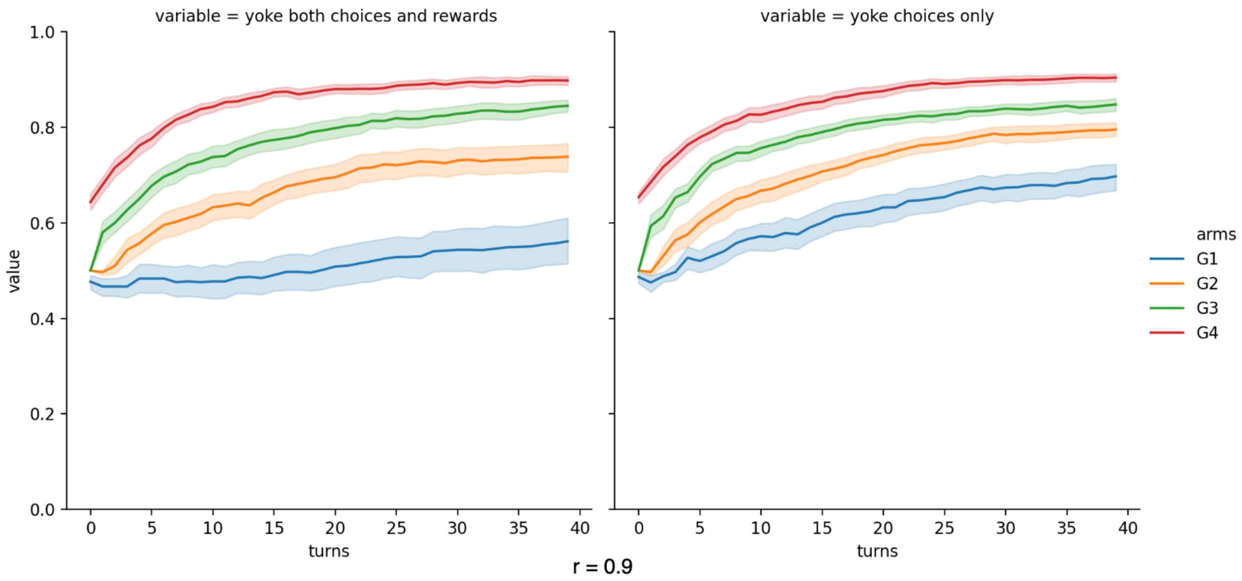
See pre-registration material for this study here (anonymized view-only for peer review).

S3.2. Human experiment pilot and pre-registrations: Pilot 2

We piloted the yoked designs before Study 2 to make sure the study materials and fetch response functions worked well, as well as to get an estimation of sample size. With the same experimental materials described above, we conducted a pilot study with $N = 50$ per condition and 4 conditions in total. See details in pre-registration materials for this study here (pilot 2) and here (Experiment 2); both are anonymized view-only for peer review.

S3.2-1a. Simulation results for yoke designs: yoke both on the left and yoke choice only on the right.

According to our rational model, we hypothesized the yoke-both choice and reward condition to mimic responses from the self-select condition because agents see exactly the same choices and rewards in the same order, which effectively is the same model. See plot on the left in Figure S3.2-1a for simulation results (identical and high rewards of .9 success probability, 40 rounds of game, and 4-arm). In contrast, we hypothesized the yoke choice only condition would shrink the estimated reward differences. Because agents might encounter different rewards and by integrating new pairs of rewards and choices, the estimated rewards should change. See plot on the right in Figure S3.2-1a for simulation results.

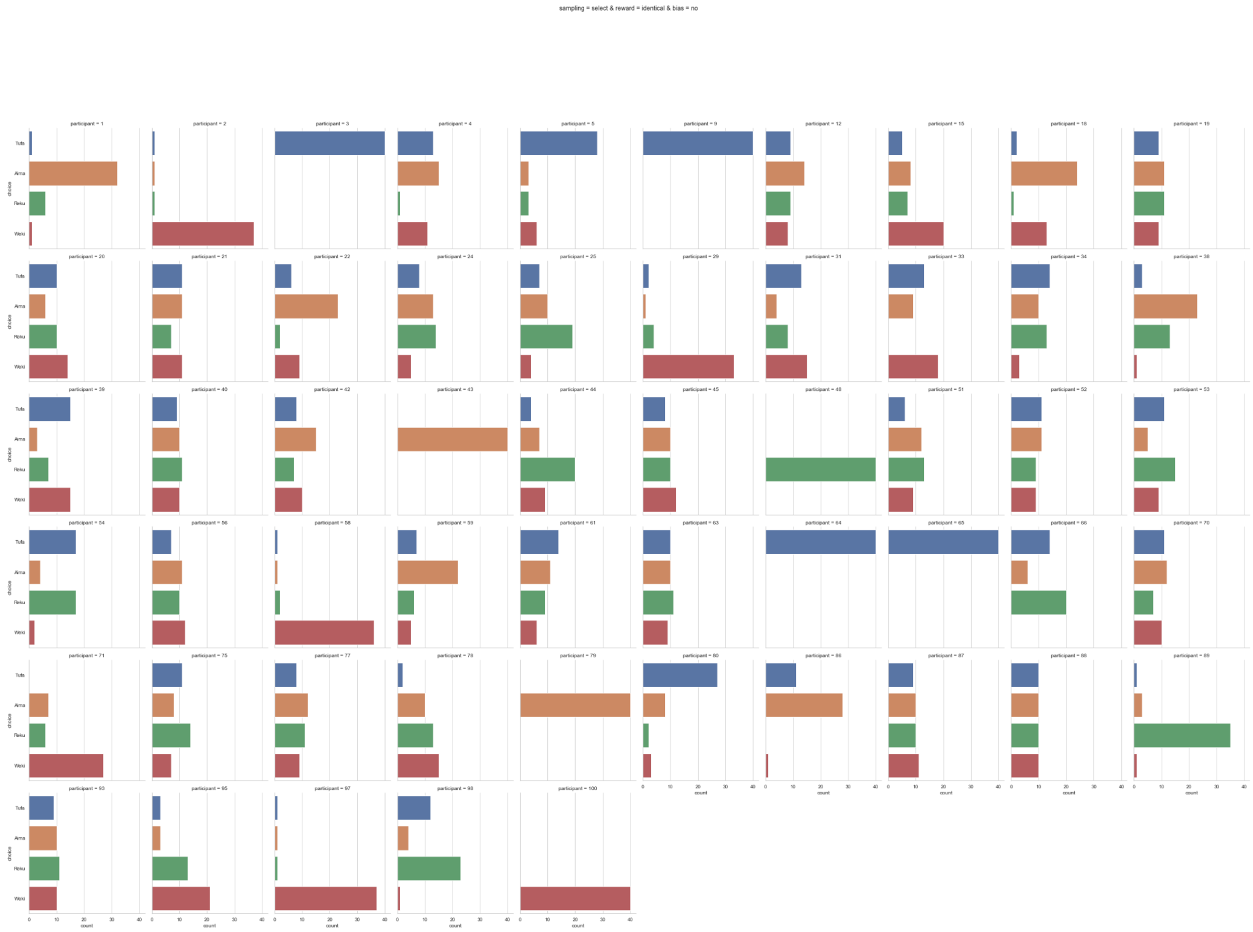


S4.1. Human experiment empirical analyses: Experiment 1

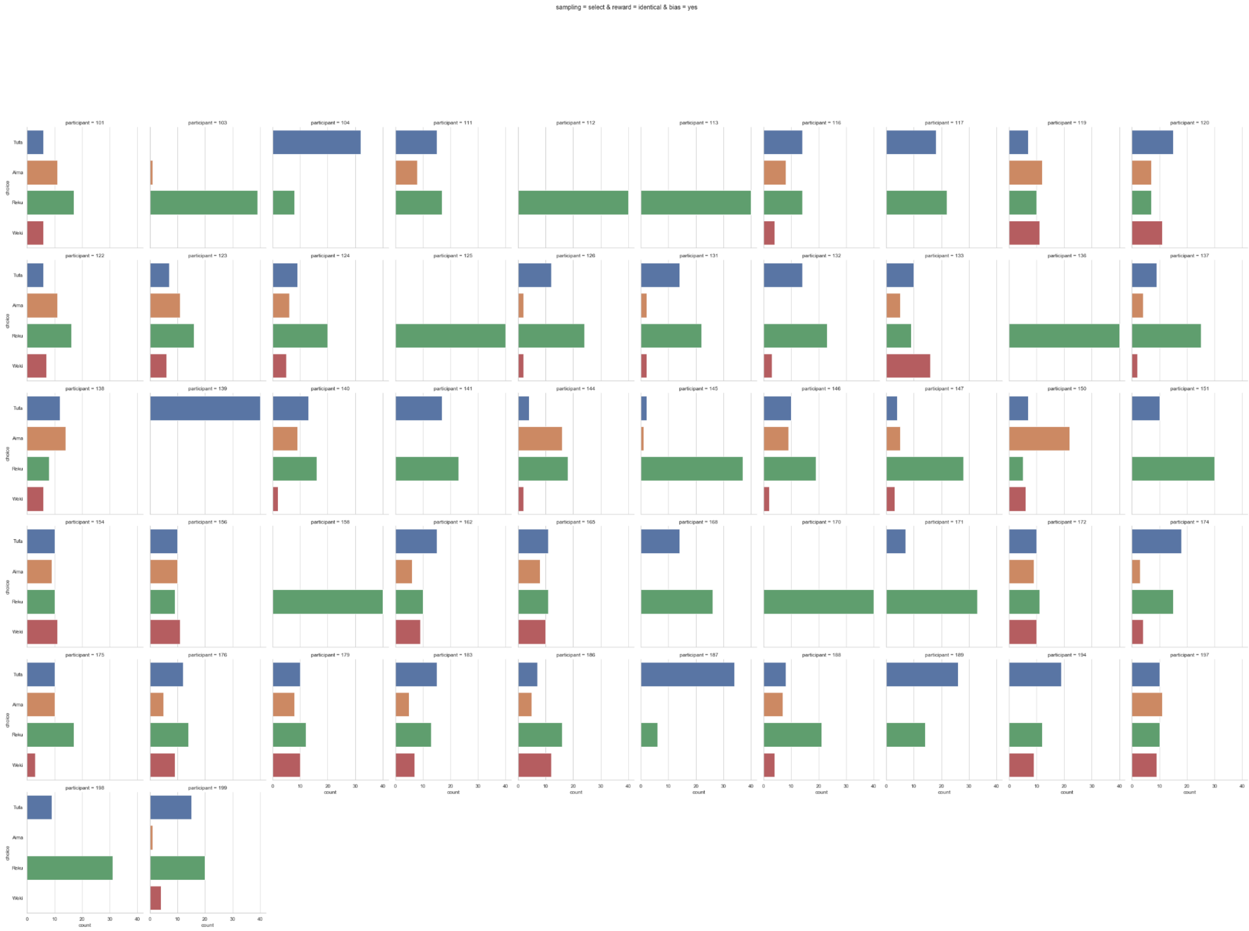
S4.1-1. Number of participants:

Reward structure	Prior bias	Sampling strategy	*N*
Different	No	random-meet	42
		self-select	58
	Yes	random-meet	44
		self-select	56
Identical	No	random-meet	45
		self-select	55
	Yes	random-meet	47
		self-select	52

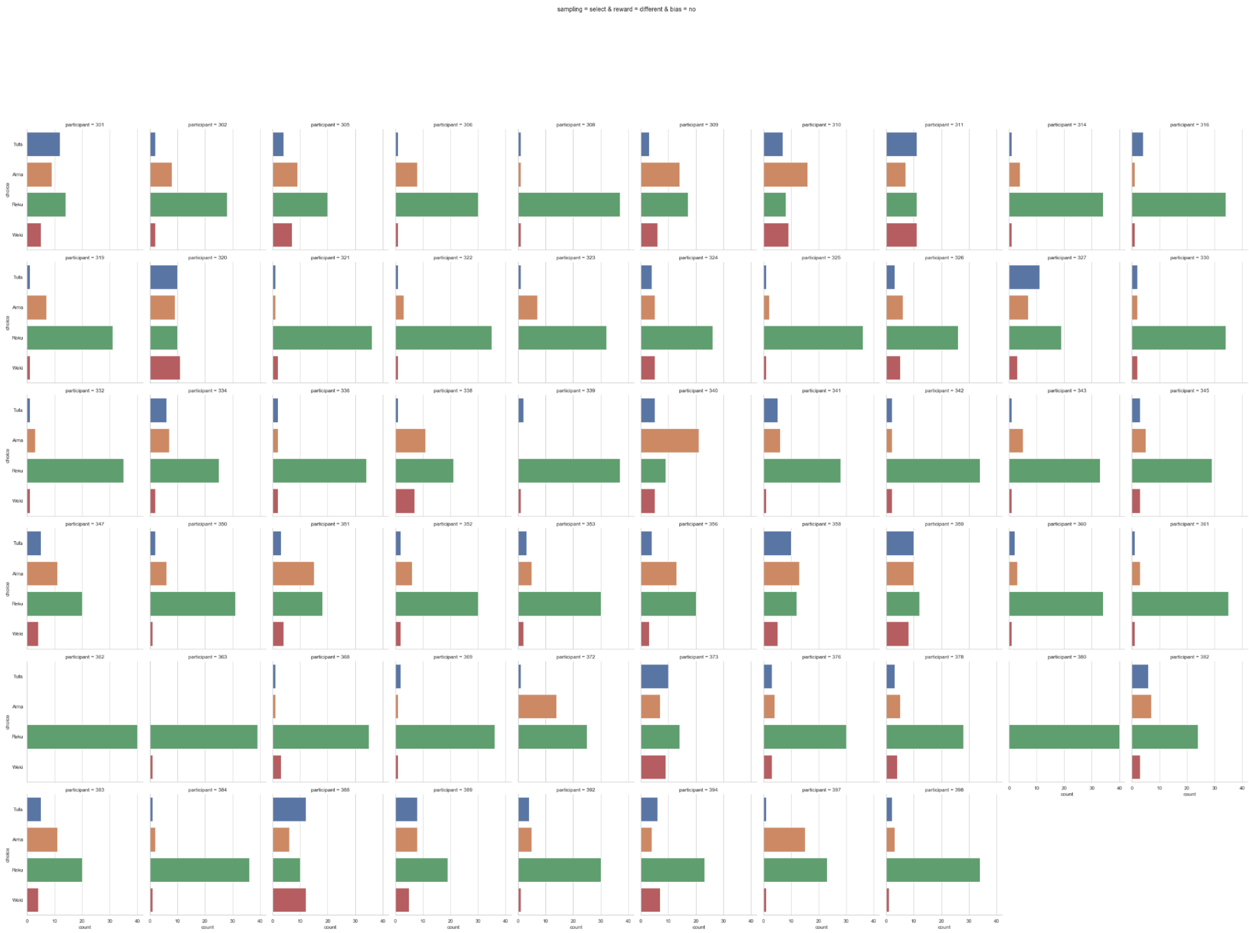
S4.1-2a. Individual-level choices in identical reward and no prior bias condition: We can see not all participants follow Thompson sampling strategies. Some follow strategies that more closely resemble dynamic programming such as participants in row #1 column #2 and row #1 column #3. Some follow random sampling strategies such as participants in row #1 column #10 or row #6 column #1. However, on average, we observe more participants follow Thompson sampling than other strategies.



S4.1-2b. Individual-level choices in identical reward and prior bias condition: Now we see many responses predominantly appearing as green bars, that is Rekus in our story. This is because participants received prior bias “Rekus are wealthy and helpful” before the game.



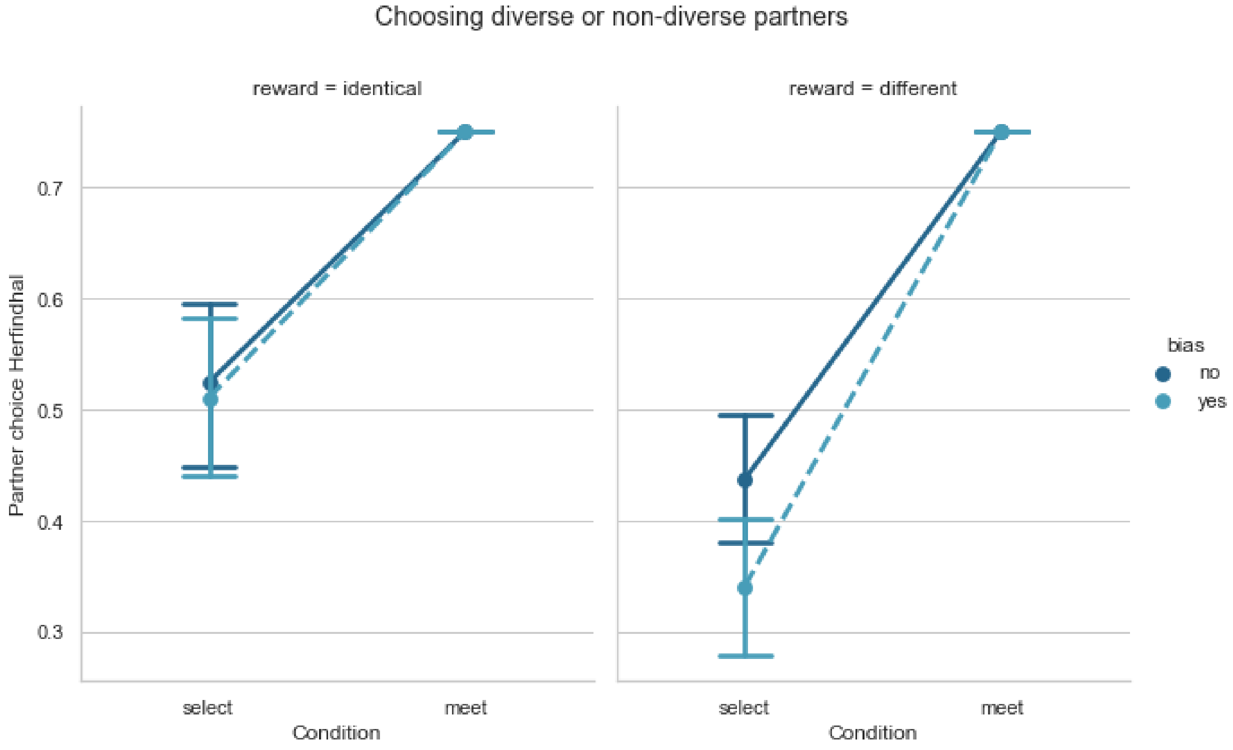
S4.1-2c. Individual-level choices in different rewards and no prior bias condition: Now we see again many responses predominantly appearing as green bars, corresponding to Rekus in our story. This time, it is because the rewards are set to be different and Rekus are set up as the most rewarding group. Participants are able to learn quickly which group is most rewarding if there are group-level differences.



S4.1-3a. Partner choice Herfindahl score: We represent the composite score as the vertical-axis and conditions as the horizontal-axis here. In both identical and reward different conditions, participants in random-meet conditions are ceiling - they meet an equal number of Toma groups, which gives a Herfindahl score of .75 (absolute diverse). In contrast, participants in the self-select condition give a much lower Herfindahl score (relatively homogenous) - they meet an unequal number of people from different Toma groups. The main text bar plot represents the raw times of interactions, but the key results are the same.

In our statistical analysis, we ran simple linear regression with condition (self-select coded as 1 v. random-meet coded as 0) as the independent variable and Herfindahl score as the dependent variable. Beta coefficients derived from the model thus indicate main condition differences, if any.

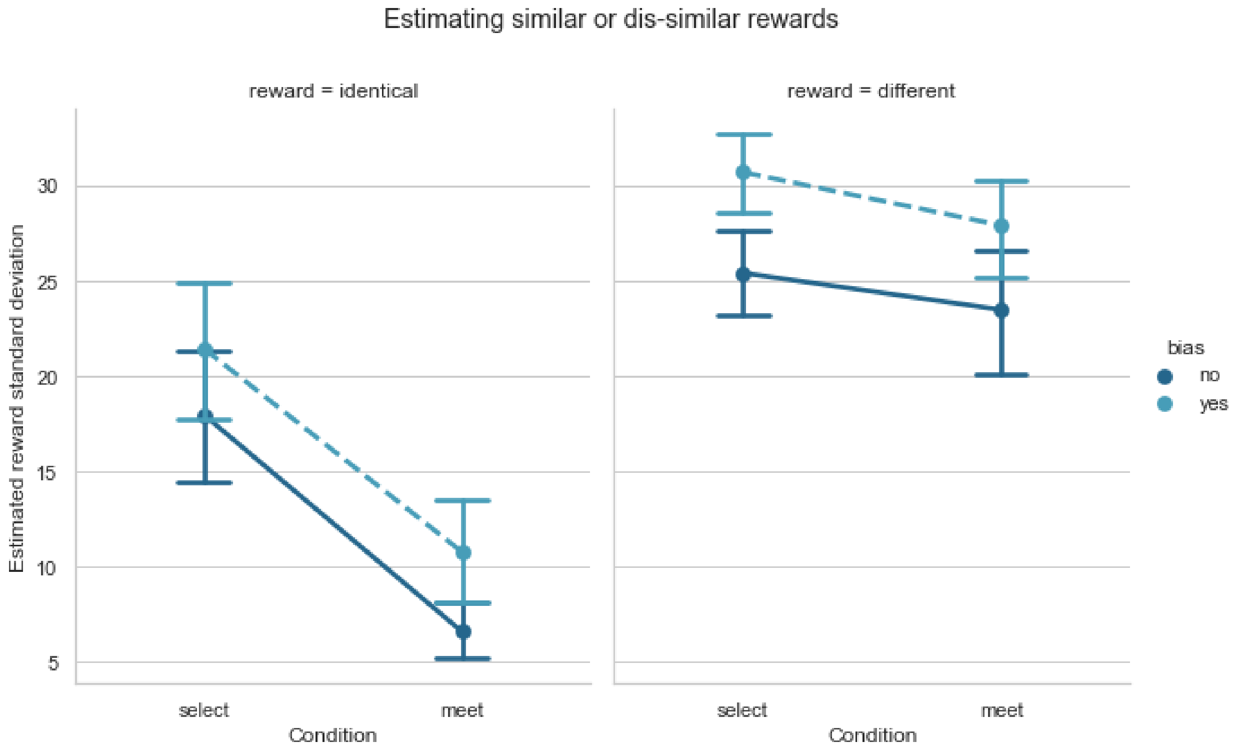
We tested this model in identical reward without prior bias (b = -.226, 95% *CI* [-.306, -.146], $p <$ .001), identical reward with prior bias (b = -.240, 95% *CI* [-.315, -.165], $p <$ .001), different reward without prior bias (b = -.313, 95% *CI* [-.379, -.247], $p <$ .001), and different reward with prior bias (b = -.409, 95% *CI* [-.479, -.340], $p <$ .001) separately. See visualizations below and R code in the online repo.



S4.1-3b. Reward estimation standard deviation: Supplementing raw scores in the main text, here we plot composite standard deviation of four groups estimated rewards on the vertical-axis and conditions on the horizontal-axis. Here we see participants in the random-meet condition, especially if they are in the reward identical environment, reported much lower standard deviations than participants in the self-select condition. We also see a main effect of prior bias and a main effect of reward structures.

For statistical analysis, we ran a similar model (independent variable: self select coded as 1 v. random meet coded as 0) but replacing partner choice Herfindhal with reward estimation standard deviation as the dependent variable. Again, beta coefficients derived from the model indicate main condition differences, if any.

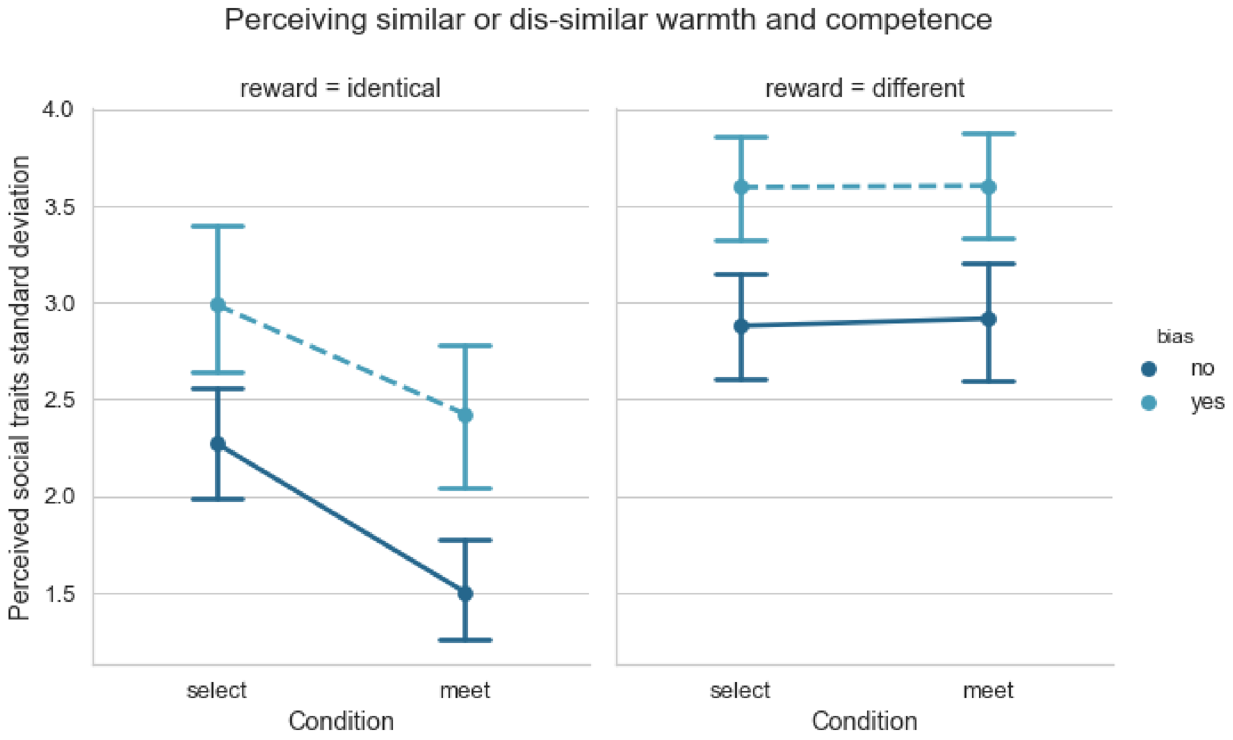
We tested this model in identical reward without prior bias (*b* = 11.354, 95% *CI* [7.387, 15.321], *p* < .001), identical reward with prior bias (*b* = 10.650, 95% *CI* [6.193, 15.108], *p* < .001), different reward without prior bias (*b* = 1.921, 95% *CI* [-1.980, 5.821], *p* = .331), and different reward with prior bias (*b* = 2.778, 95% *CI* [-.384, 5.941], *p* = .084) separately. See visualizations below and R code in the online repo.



S4.1-4. Perceived warmth and competence: As exploratory dependent variables, we collected ratings on perceived warmth and competence about each group after the game. The vertical-axis represents perceived differences between Tufa, Aima, Reku, and Weki in terms of a two-dimensional space of warmth and competence. The higher score indicates the more perceived differences and the lower score indicates the more perceived similarities. The horizontal-axis represents condition.

We observe that participants in a reward-identical environment without prior bias, those who are in the random-meet condition tend to perceive Toma groups as more similar to each other than those who are in the self-select condition. Running a similar simple regression model as before but using warmth-competence dispersion as the dependent variable and self-select v. random-meet condition as the independent variable, we found statistically significant main condition differences ($b = .768$, 95% *CI* [.366, 1.170], $p < .001$).

In short, perceived warmth and competence largely confirms the estimated reward measure. In addition, the correlation between warmth-competence dispersion and estimated reward standard deviation is moderate $r(397) = .566$, $p < .001$.



S4.2. Human experiment empirical analyses: Experiment 2

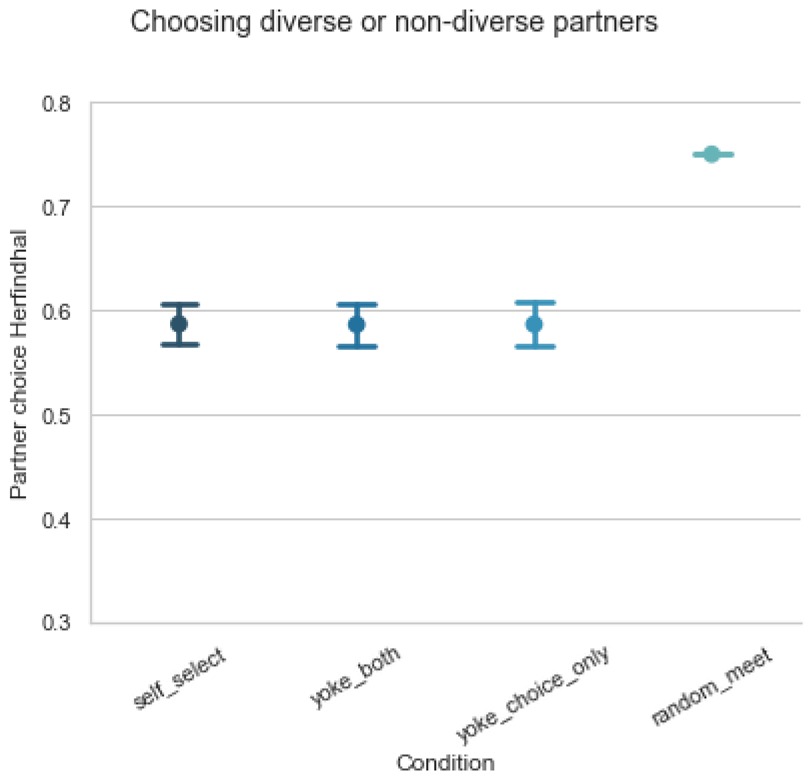
S4.2-1. Number of participants:

Reward structure	Prior bias	Sampling strategy	*N*
Identical	No	self-select	502
		yoke-both	500
		yoke-choice-only	501
		random-meet	502

S4.2-2. Individual-level choices: see online codebook due to page limits (*N* = 502 subplots). In short, we observe similar patterns as in S4.1-2a.

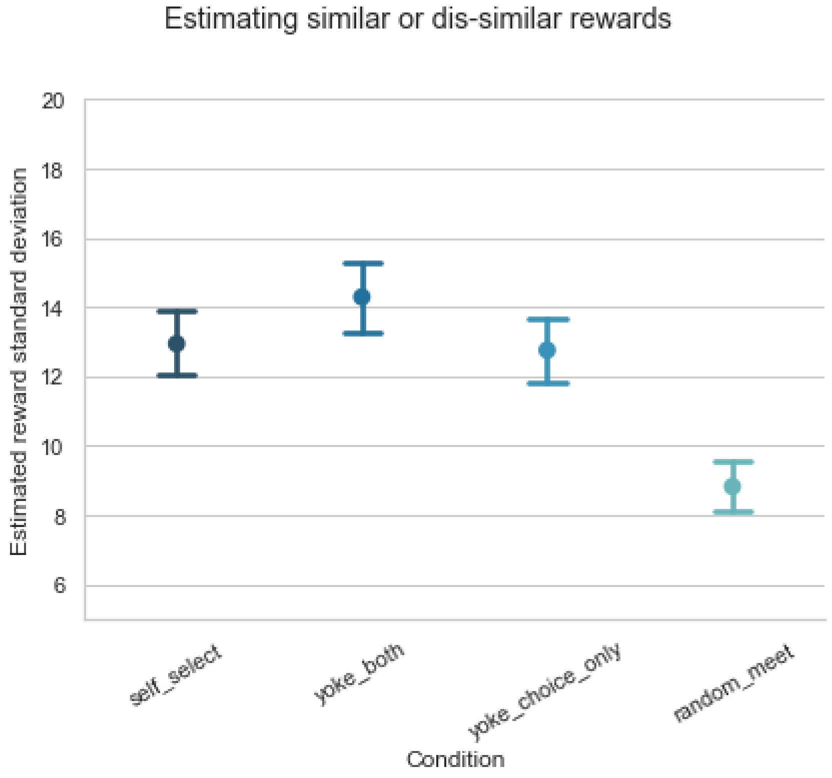
S4.2-3a. Partner choice Herfindahl score: horizontal-axis is condition, whereas vertical-axis is the composite score that indicates how diverse sample participants have seen (the higher the more diverse). As expected, participants in the random-meet condition are ceiling and self-select, yoke-both, and yoke-choice-only conditions are less likely to see diverse samples.

Similar to Experiment 1, we ran a simple multiple linear regression with self-select as the reference condition. Independent variables are thus the four experimental conditions and dependent variables are partner choice herfindahl scores. Statistics are reported in the main text, see visualizations below and R code in the online repo.



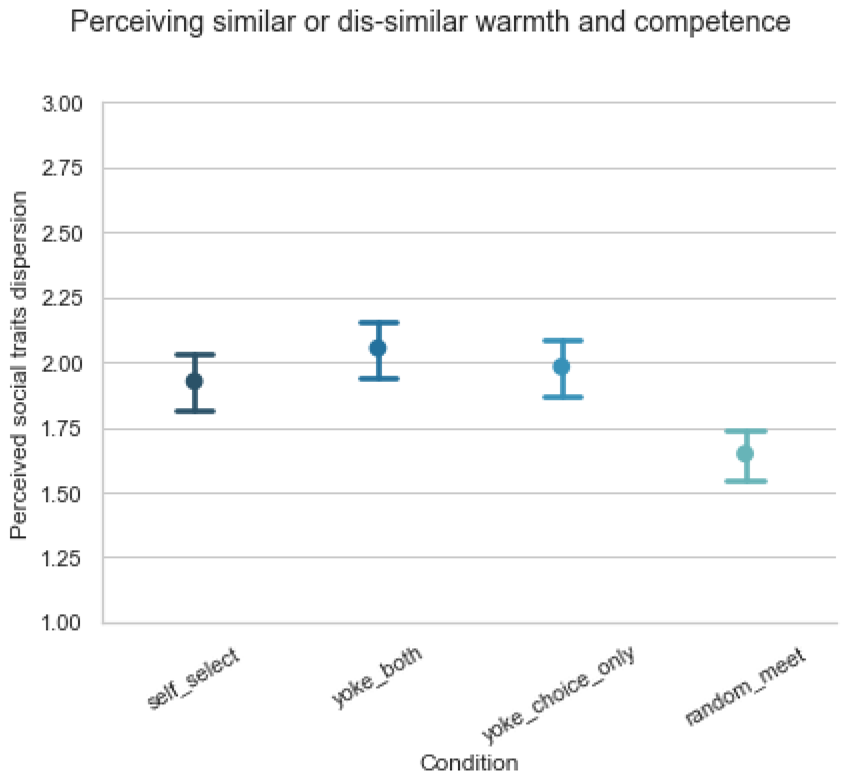
S4.2-3b. Reward estimation standard deviation: horizontal-axis is condition, whereas vertical-axis is the composite score of standard deviation of estimated rewards among the four groups; the lower the more similar. We ran a similar analysis but replacing partner choice Herfindhal with estimated reward standard deviation in this study.

As expected, participants in the random-meet condition are more likely to think the four groups are similar to each other as compared to the other three conditions. Participants in the yoke-both condition are slightly more likely to think the four groups are different than the self-select condition. Participants in the yoke-choice-only condition are slightly less likely to think the four groups are different than the self-select or yoke-both condition. However, note the effect sizes between self-select, yoke-both, and yoke-choice-only are small.



S4.2-4. Perceived warmth and competence: Repeating our exploratory analysis with perceived warmth and competence, we found smaller warmth and competence differences in the random-meet condition than the other three conditions. A simple linear regression compared random-meet against self-select condition found a statistically significant condition difference (*b* = .277, 95% *CI* [.427, .127], *p* < .001). To a smaller extent, perceived warmth and competence differences replicate the trends of estimated rewards: Participants in the yoke-both (*b* = .128, 95% *CI* [.427, .127], *p* =.096) think groups are slightly more different than participants in the self-select condition, but participants in the yoke-choice-only (*b* = .056, 95% *CI* [.427, .127], *p* =.464) condition do not think groups differ that much in terms of warmth and competence than participants in the self-select condition.

In short, although estimated reward standard deviation and perceived warmth and competence dispersion positively and moderately correlated with each other (*r*(2003) = .553, *p* < .001), in this study, we found the effects to be smaller with warmth and competence perceptions than reward estimates. We note that warmth and competence perceptions can be related to distinct types of tasks, making the current task of starting a business too broad to make concrete predictions about specific stereotype contents. See more in the Discussion in the main text.



S5.1. Discussion of illusory correlation

Chapman (1967) introduced the concept of illusory correlation as “the report by observers of a correlation between two classes of events which, in reality, (a) are not correlated, or (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from which is reported.” He also suggested “prejudice is in part an illusory correlation between race and negative traits while halo effect is an illusory correlation between race and positive traits.” In a series of pair-of-words experiment, participants estimated long words to co-occur more (estimated to have appeared more than 40% in the training trial) than they actually were (33%). Therefore, he proposed the co-occurence of distinctive stimuli can result in an overestimation of the frequency with which the two events occur together.

Hamilton & Gifford (1976) investigated (b) for social perception. The mechanism underlying illusory correlation is explained as “observers are more attentive to distinctive than nondistinctive stimuli, and the heightened attention to a distinctive stimulus should result in a greater encoding of that information. Extending this line of reasoning, the co-occurence of two distinctive events should be particularly salient to an observer, resulting in increased attention to and more effective encoding of the fact that the two events occur together, thereby increasing the subjective belief that a relationship exists between them.” In this study, the pair contains novel minority v. majority human group and desirable v. undesirable traits. Illusory correlation was measured as recall if A or B had performed each of the behaviors in the training trial. They found participants over-recall undesirable-minority pairs.

Sherman, Kruschke, Sherman, Percy, Petrocelli, & Conrey (2009) integrated exaggerated stereotypes and illusory stereotypes with attention theory of category learning mechanism. According to the model, the features of majority groups are learned earlier than the features of minority groups. In turn, the features that become associated with a minority are those that most distinguish it from the majority. This second process is driven by an attention-shifting mechanism that directs attention toward group-attribute pairings that facilitate differentiation of the two groups and may lead to the formation of stronger minority stereotypes. See also ***Alves, Koch, & Unkelbach (2018)***.

Costello & Watts (2019) proposed a rational analysis of illusory correlation using the rule of succession. They argue that patterns in illusory correlation are consistent with the rule of succession (p = k+1/n+2; k being occurence and n being events). Judgements are regressive (closer to 0.5) compared to observed sample proportions, with this regression being stronger for judgements about the minority than the majority. The assumption is differential frequency and inductive statistical inference.

Our work differs from the above works in that we

(a) **Operationalize stereotypes in straightforward and abstract terms (i.e., perceived differences among groups).** Rather than comparing recall of traits or proportion of trait estimates for each group separately, we directly measure the concept of perceived differences among groups, by using standard deviation. Admittedly, our operationalization is more abstract than prior work. We do not have any concrete traits, e.g., nice, intelligent, but we use expected utility to subsume specific traits. Future work will test this with more concrete traits.

(b) **Minimize the assumption of unequal group size (i.e., information deficits in the abstract).** All illusory correlations in the above studies start with the assumption that group sizes differ -- a majority group and a minority group. In contrast, our proposal does not require group size differences. We can have an equal number of members and an equal ratio. Moreover, we provide a procedural explanation for why people end up seeing different group sizes -- because of the process of exploratory sampling. Furthermore, we provide a possible explanation on why people often see majority group size correlate with positive attributes -- because of the self-interest maximizing goal. In sum, our proposal explains some precedent conditions for illusory correlation mentioned in earlier works.

(c) **Minimize the assumption of cognitive attention (i.e., cognitive limitations in the abstract).** Except work by Costello & Watts, or Fiedler et al., other primary mechanisms regarding illusory correlation center around selective attention, such as Sherman et al.'s attentional shift mechanism. We do not dispute such a possibility, but our work suggests even without attentional shifts, we still see illusory group differences emerge. Throughout our experiments, we only need to causally manipulate sampling strategies to see an emergence of inaccurate impressions (i.e., Experiment 1). One way to test how the attention mechanism works in our experiment would be to manipulate participants' attention (block attentional shifts vs. encourage attentional shifts), which should be a future direction.

(d) **Describe why rather than how (i.e., functional in the abstract).** We aim to provide an abstract functional account of why we see inaccurate stereotypes rather than a concrete psychological process. Exploratory sampling describes one possible mechanism that produces inaccurate stereotypes, and is a consequence of seeking to maximize long-term rewards. Unexpectedly, such an adaptive solution (for the goal of self-interest maximization) produces suboptimal impressions (for an alternative goal of impression formation). Hence, our paradigm aims to point out one overlooked relationship between individual utility-maximizing choices and societal collateral damage (see more in the General Discussion).

S5.2. Functional analysis framework for social stereotyping

Our analysis of social stereotypes as emerging from local exploration is in the spirit of Anderson's (1991) idea that psychological phenomena can be explained in terms of functional solutions to problems that arise in people's environment. In contrast to mechanistic explanations -- such as internal structures or cognitive processes -- a functional analysis focuses on purposive explanations. It asks: what are the goals of the system; what are the nature of the environments; and what optimal solutions can be derived to achieve the goal in that environment. As such, a functional analysis can potentially explain why a phenomenon exists: because it approximates an optimal solution to attaining some adaptively relevant goal.

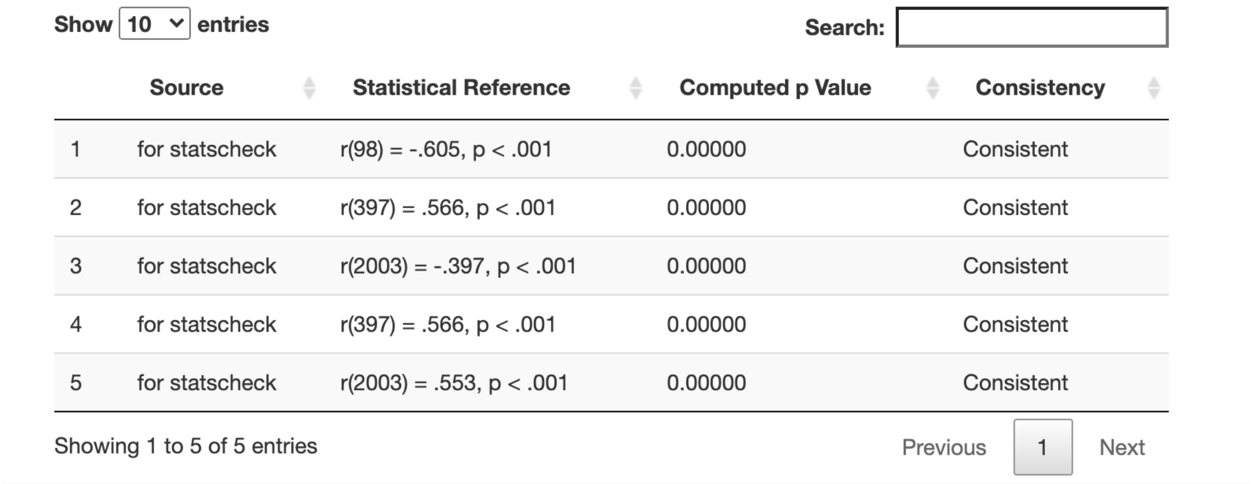
Within our context, the problem of social stereotyping can be formulated as the following:

- The goal is to maximize one's cumulative long-term rewards.
- The environment poses identically high rewards to each of the candidate groups.
- The constraint is the sequential experience, that one can only learn potential rewards from interacting with one member from one group sequentially.
- One optimal solution to this problem under infinite interactions is Thompson sampling.

As expected, agents can maximize their long-term rewards through such exploratory sampling (i.e., earn many if not all points by playing with any of the groups). Hence, exploratory sampling is an optimal solution. This is also evident from the different-reward condition (i.e., the optimal strategy is to interact more with the group that has the highest expected utility, which is indeed reflected from our model simulation and human participants; Fig 2c-2d & Fig 3c-3d).

However, one unintended consequence of such exploratory sampling is inaccurate impressions about other under-explored groups. For each individual agent, the byproduct of inaccurate impressions does not matter, as it does not prohibit them from earning rewards. However for society at large, such byproducts can be detrimental, as they create illusory perceived differences among groups.

S5.3. StatsCheck Reports



	Source	Statistical Reference	Computed p Value	Consistency
1	for statscheck	r(98) = -.605, p < .001	0.00000	Consistent
2	for statscheck	r(397) = .566, p < .001	0.00000	Consistent
3	for statscheck	r(2003) = -.397, p < .001	0.00000	Consistent
4	for statscheck	r(397) = .566, p < .001	0.00000	Consistent
5	for statscheck	r(2003) = .553, p < .001	0.00000	Consistent

Showing 1 to 5 of 5 entries

Note that not all statistics in the texts were detected or checked by the system. For more information, please check out our open source; url provided on the title page.