Multidimensional Stereotypes Emerge Spontaneously When Exploration is Costly

Xuechunzi Bai♣, Thomas L. Griffiths♣♦, Susan T. Fiske♣♠

♣ Department of Psychology

♦ Department of Computer Science

♠ School of Public and International Affairs

Princeton University

Multidimensional Stereotypes Emerge Spontaneously When Exploration is Costly

Word Count: 4474 words (including abstract and main text)

Project Repository: https://osf.io/6p8wu/?view_only=22709d2fd3164a90880af4b0e2679f7f (see for_review.zip for data, analysis code, study materials, pre-registration including power analysis)

Abstract

Stereotypes of social groups have a canonical multidimensional structure, reflecting the extent to which groups are considered competent and trustworthy. Traditional explanations for stereotypes – group motives, cognitive biases, minority/majority environments, or real-group differences – assume that they result from deficits in humans or their environments. A recently-proposed alternative explanation – that stereotypes can emerge when exploration is costly – posits that even optimal decision-makers in an ideal environment can inadvertently create incorrect impressions. However, existing theories fail to explain the multidimensionality of stereotypes. We show that multidimensional stratification and the associated stereotypes can result from *feature-based* exploration: when individuals make self-interested decisions based on past experiences in an environment where exploring new options carries an implicit cost, and when these options share similar attributes, they are more likely to separate groups along multiple dimensions. We formalize this theory via the contextual multi-armed bandit problem, use the resulting model to generate testable predictions, and evaluate those predictions against human behavior. In particular, we evaluate this process in incentivized decisions involving as many as 20 real jobs, and successfully recover the classic warmth-by-competence stereotype space. Further experiments show that intervening on the cost of exploration effectively mitigates bias, further demonstrating that exploration cost *per se* is the operating variable. Future diversity interventions may consider how to reduce exploration cost, such as introducing bonus rewards for diverse hires, assessing candidates using challenging tasks, and randomly making some groups unavailable for selection.

Keywords: Stereotype, Multidimension, Explore-Exploit, Generalization, Intervention

Public Significance Statements

Stereotypes are multidimensional, including features that go beyond sheer good-bad valence. Current psychological theories, which focus on social, cognitive, and sample biases do not explain the origins of such complex stereotypes. Here we show that a novel psychological mechanism can reproduce the multidimensional stratification of social groups and the resulting complex stereotypes: when individuals make self-interested decisions based on past experiences in an environment where exploring new options carries an implicit cost, and when options share similar attributes, they are more likely to separate groups along multiple dimensions. A further set of intervention experiments provides causal evidence that reducing exploration cost can substantially mitigate even complex stereotypes.

Main Text

Introduction

Social stereotypes seem to be a fundamental part of human societies. They organize expectations about gender, race, nationality, and appearance, and carry associations about perceived trustworthiness and competence (Bai et al., 2020; Bian et al., 2017; Katz & Braly, 1933; Todorov et al., 2015). People often learn these complex stereotypes from segregated societal structures signaling, for example, social status and cooperative intent (Fiske et al., 2002; Koenig & Eagly, 2014). What position a specific group occupies in such structures depends on complex economic, cultural, historical, and political circumstances. However, the mechanisms that differentiate groups follow basic psychological principles. Incorrect impressions of the abilities of different groups can emerge purely because of individuals who make social decisions facing an implicit cost for exploring new options (Bai et al., 2022). Here, we use a combination of computational simulations and incentivized behavioral experiments to show that the same mechanism can produce multidimensional stereotypes that recapitulate the axes along which people represent real social groups: differentiated stereotypes emerge spontaneously when exploration is costly and is guided by socially constructed features.

Existing psychological explanations for social stratification between groups have focused on four causes: biased decision-makers, particularly those who are high status and powerful, assigning minorities to disadvantageous positions in order to protect their ingroup or to oppress outgroups (Altemeyer, 1983; Brewer, 1999; Jost & Banaji, 1994; Pratto et al., 1994); cognitively limited decision-makers having distorted mental representations due to inherent constraints such as memory capacity or attention selectivity (Fiske & Taylor, 1984; Hamilton & Gifford, 1976; Macrae et al., 1994; Sherman et al., 2000; Trope & Thompson, 1997); statistically unsophisticated decision-makers not taking into account that they are observing unrepresentative samples, producing biases (Fiedler, 2000; Denrell, 2005; Payne et al., 2017); and, most controversially, actual group differences resulting in groups being sorted into different positions (Eagly & Steffen, 1984; McCauley et al., 1995). These four explanations thus attribute the origins of stereotypes to a defect in human decision-making or in the environmental samples.

Contrary to these notions, a recently-proposed fifth perspective informed by work in computer science that highlights the inherent tradeoff between "exploring" new options and "exploiting" existing knowledge (Sutton & Barto, 2018) posits that even optimal decision-makers might inadvertently produce bias when exploring unfamiliar options entails an implicit cost (Bai et al., 2022). While these five accounts might explain why people differentiate between groups, particularly identifying an in-group as good and an out-group as less good, they do not explain more complex stereotypes that go beyond a simple good-bad dichotomy (Abele et al., 2021; Koch et al., 2016; Nicolas et al., 2022; Zou & Cheryan, 2017). For example, stereotypes of immigrants in the US are not merely binary; perceptions vary in a multifaceted manner: Russians are seen as competent but untrustworthy, Mexicans are neither competent nor trustworthy, Native Americans as friendly but not competent, and Canadians are capable and friendly (Bai et al., 2020; Lee & Fiske, 2016). We build on the explore-exploit framework to show that multidimensional social stratification need not to be rooted in flaws in humans or the environments, it is simply a consequence of the way social decisions are often posed. Costly exploration, combined with socially constructed features that provide a basis for generalization, is sufficient to produce rich multidimensional stereotypes.

To illustrate our proposed mechanism and to anticipate the methods used in our experiments, imagine a manager hiring individuals from different social groups for different jobs

(Figure. 1). The manager's goal is to ensure successful outcomes in these jobs. Assume individuals from all groups are equally and highly likely to succeed in all kinds of jobs. The manager does not know this and seeks to learn how well the different groups perform based on experience. Unfortunately, the learning process suffers from a serious constraint: The manager can only observe the performance of people they hire, so they remain ignorant of how well the people they did not hire could have done.

As a specific example, consider five jobs that vary on two features: high-status and high-trust doctors and veterinarians; high-status and low-trust lawyers; low-status and high-trust childcare aides; low-status and low-trust garbage collectors (Fiske & Dupree, 2014; Koenig & Eagly, 2014). As jobs become available one by one, the manager assigns people from different groups to do each job in turn and observes the performance of the hired individuals.

Initially, when a garbage collector position opens, the manager may randomly choose a person from one group (blue) without enough information to make a better decision. But they learn that it is a good choice. Next, the manager must choose somebody for a doctor position, still without enough evidence to support a definitive decision, so perhaps they want to stick to the same group one more time but quickly discover it is a poor decision (suppose they happen to hit the rare incompetent individual in this population where most groups can do most jobs). A third job, a veterinarian position, is available. Although the manager has not hired a veterinarian before, given that veterinarians share similar features with doctors, managers may generalize from their past experiences. Given their past negative experience with blues as doctors, the manager may switch to a different group (yellow). They learn that the newly recommended individual performs well. The process continues.

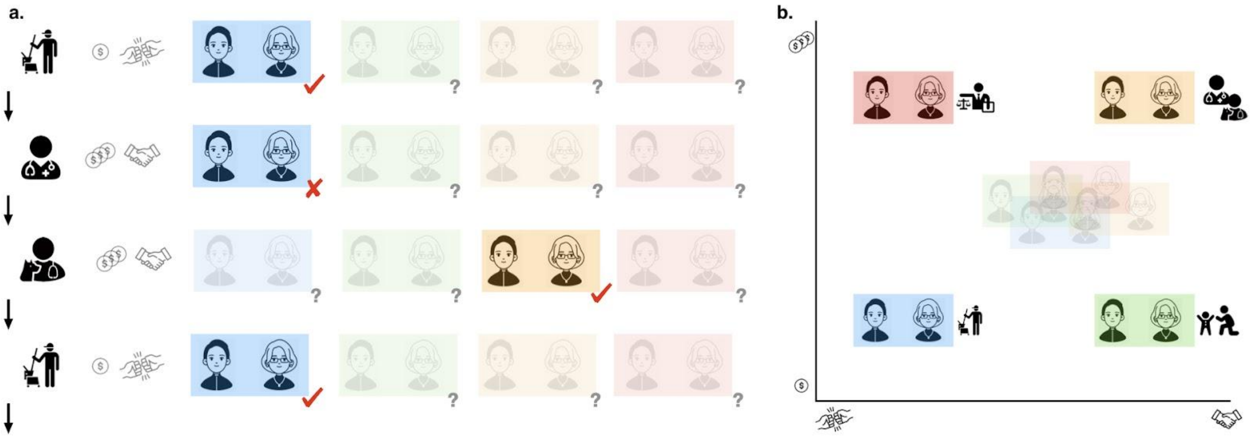


Figure. 1. The hiring task as a contextual multi-armed bandit. An example illustrates how making new decisions based on past (selective) experiences can create a stratified unit that produces multi-dimensional stereotypes that are incorrect. Panel a shows example jobs, their associated features such as social status and cooperative intent, and four groups. Each decision only has one group being hired, whose performance is then revealed and is used to guide new decisions, while the other three groups remain unknown. Panel b shows mental representations after these decisions are made. The example mental map is organized by two features – competence and trustworthiness. The true situation is pictured in the background, while the incorrect impressions of the groups formed by the decision-maker are shown in the foreground.

Remember, the underlying probability of being successful is identical and high for all pairs of jobs and groups. Despite individual variation, on average, every group is just as good as any other group at performing all jobs. Intuitively, initial positive experiences recommending members of one group for garbage collectors may encourage the manager to recommend more members from that group as garbage collectors or for similar jobs. Consequently, the manager is less likely to recommend people from other groups for the same positions, or people from that group for other jobs. If so, the manager has introduced social stratification, hiring more people from one group for low-status and low-trust jobs. Observing this pattern, the manager and others might wrongly conclude that the overrepresented group in these positions is incompetent and untrustworthy.

This example illustrates how a series of seemingly adaptive decisions can produce a social reality that sorts members of different groups into distinct positions, without needing to appeal to group motives, cognitive limits, sample imbalances, or group differences. This behavior is adaptive for the individual decision-maker as it optimizes hiring performance in two key ways. First, it minimizes the implicit cost from exploring a new uncertain group, which might not perform as reliably as a more familiar choice (Bai et al., 2022). Second, it further reduces the exploration cost by generalizing shared features across positions. Using these features, the decision-maker can recommend similar but not identical positions to the same group (Shepard, 1987). Despite multiple adaptive benefits to the individual, this behavior is detrimental to society because the byproduct of these decisions is a biased and stratified representation of reality. Not only do some groups receive inadequate exploration, but the underlying features associated with them also become the foundation for complex, multidimensional stereotypes. Multidimensional stratification emerges from adaptive individual decisions for the individual, but decisions that are maladaptive for the collective.

This minimal explanation for the origin of stereotypes is challenging to test because multiple mechanisms are confounded in studies of stereotypes based on real-world knowledge. To address this challenge, we used a combination of computational modeling and incentivized behavioral experiments. The computational model precisely defines the problem being solved and demonstrates the emergence of stereotypes in the absence of group motivations, cognitive limitations, unequal sample size, or differing group qualities (see below, **Model**). The behavioral experiment enriches the simple scenario assumed in the model with as many as 20 real-world jobs. Both computational agents and human participants stratify their environments and form stereotypes, even along multiple dimensions, simply because feature-based exploration has intrinsic costs (see below, **Experiment**). Intervening to reduce these costs however reduces stratification and stereotypes (see below, **Evaluating Interventions**).

Results

Model. To formalize our hiring problem, we adapt the contextual multiarmed bandit task – a fundamental problem explored in theoretical treatments of sequential decision-making and reinforcement learning in computer science and related disciplines (Sutton & Barto, 2018). In a multiarmed bandit task, an agent chooses actions (pulling an "arm" of the "bandit," an old-fashioned gambling machine) to receive rewards over multiple rounds. Each arm has a probability distribution over rewards. In each round, the agent selects an arm and receives a reward sampled with the corresponding probability. The agent wants to maximize their cumulative rewards but is unaware of the reward distributions associated with the arms. The agent thus needs to balance two competing options: *exploring* a new arm to learn its reward, and *exploiting* the arm that is known to give the highest expected reward.

Many real decisions involve choosing between options that are differentiated by observable features. The *contextual* multiarmed bandit task captures this by assuming that the reward distribution depends not only on the arm but also on a set of features that describe the decision context on that round (Li et al., 2010). Instead of estimating the reward distribution for each arm, the agent now estimates the function that maps contextual features to reward distributions. While this problem is harder to solve than the simple multiarmed bandit, it yields greater flexibility, as the agent can learn to generalize to future similar but not identical situations based on their features. This is the critical modification that makes multidimensional stereotypes emerge.

While there are no known optimal solutions for the contextual bandit task, we use a Bayesian approach called Thompson sampling (Thompson, 1933; Agrawal & Goyal, 2012). Thompson sampling uses Bayesian inference to estimate the probability of reward associated with each arm, and then samples an arm with a probability that matches the posterior probability of that arm offering the best chance of reward. This approach has been shown to be an effective model of human choices and social interactions (Schulz et al., 2018; Bai et al., 2022). To learn the function between contextual features and reward distributions, we employ Bayesian logistic regression (Li et al., 2010; Chapelle & Li, 2011).

Using the described model, we simulated the behavior of adaptive-decision agents who follow Thompson sampling and random-decision agents who do not maximize rewards or use past experiences in choosing among four groups over 40 choice trials (see *SI* for model details). The choices involved allocating members of the different groups to jobs, where each job had a known set of features reflecting the need for trustworthiness and competence, and the adaptive-decision agents' estimated parameters for each group indicating the extent to which they had these features. The underlying rate at which rewards were delivered to all groups was the same: rewards were sampled from a Bernoulli distribution where each individual had a 90% chance of succeeding in the job, hence delivering a reward for the decision-maker. Note that the simulated agents are initialized with an uninformative prior follows a unit normal distribution for each group. The agents do not have parameters for group motivation or memory limitations, and the ground truth dataset does not contain unequal population sizes or different reward probabilities (see other simulation variants such as differing ground truth and differing prior beliefs in *SI*).

Nonetheless, the simulation reveals that adaptive-decision agents, while attempting to maximize rewards through past experience, are more likely to allocate groups differentially and form stereotypes compared to random-decision agents. We illustrate this using an ordinary-least-squares linear regression model with the agent type as the predictor variable (adaptive coded as 1

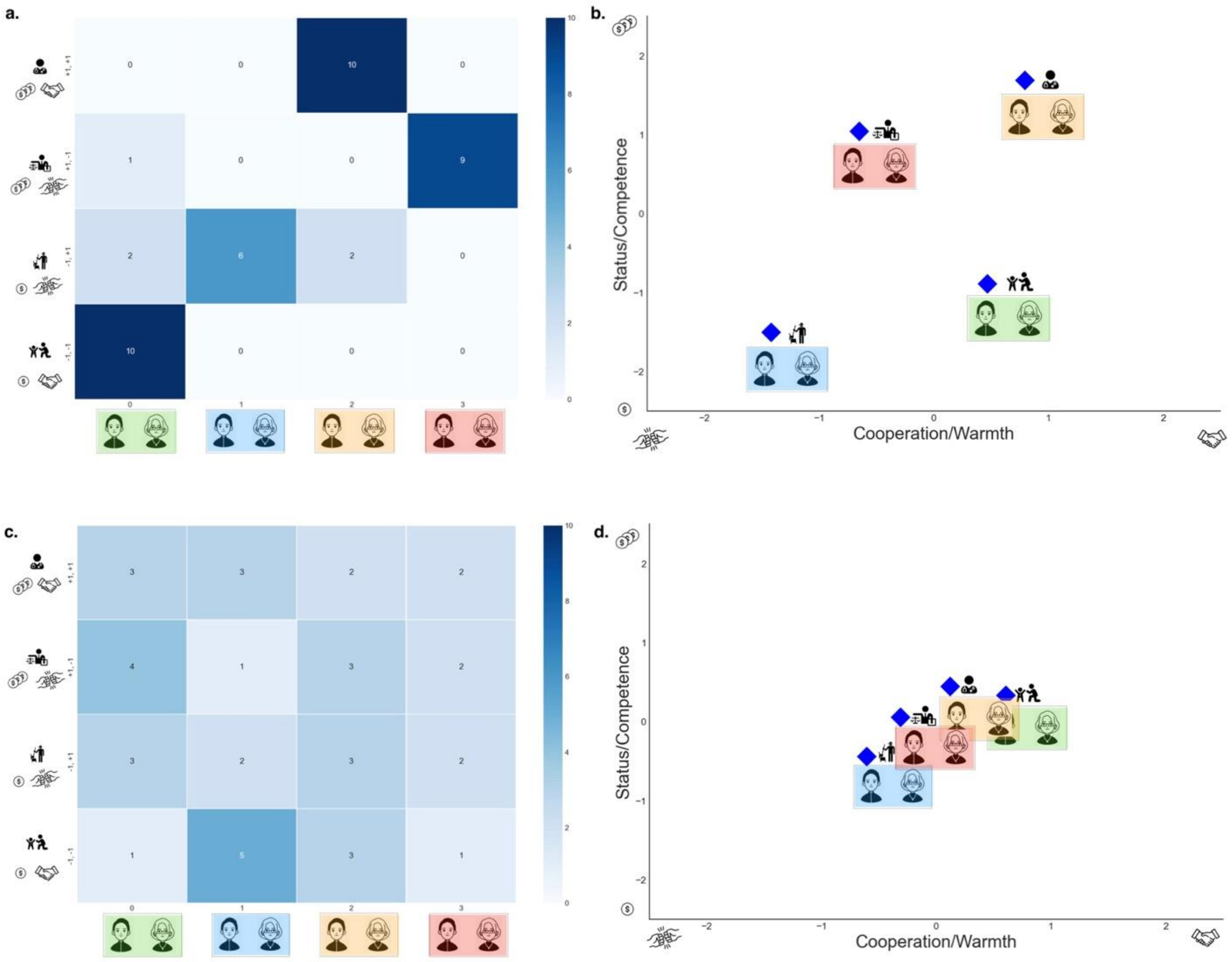


Figure. 2. Two example simulated results from an agent who makes adaptive decisions (in panels a-b) and another agent who makes decisions at random (in panels c-d). The heatmaps on the left panels show how many times a group, on the horizontal axis, is recommended for a job, on the vertical axis. The scatterplots on the right panels show estimated coefficients for the four groups on the two binary features. For aggregate simulation results see *SI* simulation section.

vs. random coded as 0) and the entropy of the distribution of choices over groups (i.e., choice entropy) and the distance between estimated parameters for the groups (i.e., stereotype dispersion) as the outcome variables. This model shows that the adaptive-decision agents show a lower entropy, indicative of stratified choices (b = -.645, 95% *CI* [-.614, -.676], p < .001), and a bigger distance, indicative of differentiated stereotypes (b = 1.447, 95% *CI* [1.596, 1.297], p < .001; Figure. 2 for prototypes) as compared to the random-decision agents. Stratified choices and dispersed estimated parameters emerge from the agents trying to solve the explore-exploit dilemma to maximize their rewards, while minimizing the hidden cost of exploring the unknown.

Experiment. We tested the predictions of this model in a large-scale online experiment in which participants (*N* = 1310) made hiring decisions involving novel social groups. Participants were told that they had been recruited by the mayor of a made-up place, Toma City, to recommend members of four groups of people, the Tufas, Aimas, Rekus, and Wekis, for different jobs. The better recommendations the participants make, the more money they earn. To test whether participants generalize their experiences from a few limited jobs to a large amount of similar but not identical jobs, we prepared 20 different kinds of jobs (Dupree & Fiske, 2014; Koenig & Eagly, 2014; see *SI* for a preliminary study norming these jobs), and jobs open one at a time at random. In the adaptive exploration condition, participants make decisions sequentially and learn the outcome of their recommendation after each decision, earning 1 point or 0 points. In the random exploration condition, participants observe the mayor making random decisions. This minimal design aimed to reduce the impact of group motivations, cognitive limitations, unrepresentative sampling, and quality differences, while focusing on the causal effects of adaptive versus random exploration (see *SI* for experimental designs).

Confirming the model predictions, the human data show statistically significant differences in choice entropy between the adaptive exploration condition and the random exploration condition (*b* = -.476, 95% *CI* [-.437, -.514], *p* < .001). This analysis controls for individual differences in age, gender, race, education, and political orientation. Participants who make their own decisions display lower entropy, corresponding to more stratified and unequally distributed choices (Figure. 4a. “Default”). In contrast, participants who observe random decisions from the mayor display higher entropy with less stratified and more equally distributed choices (Figure. 4a. “Ideal”). Moreover, compared to participants who observe random decisions, participants who adaptively explore are more likely to report larger mental distances in the trustworthiness-competence space (*b* = .343, 95% *CI* [.597, .089], *p* < .001; Figure. 4b. “Default” versus “Ideal”; Figure. 3 for example participants). The stratified choice also holds for imagined future hires where participants make new decisions regarding unseen applicants. The stereotype dispersion also holds for status and cooperation dimensions, which are theorized as structural antecedents of competence and trustworthiness (Abele et al., 2021; Fiske et al., 2002; see *SI* for more results).

Two results are worth highlighting: First, we see evidence for the emergence of multidimensional stereotypes. As shown in Figure. 3b, participants do not simply polarize Toma groups as the uniformly good versus the utterly bad ones. Rather, they clearly differentiate along at least two dimensions – for example, Tufas are competent but not trustworthy or Wekis are incompetent but trustworthy (Figure. 3a). Second, we see evidence for generalization (Shepard, 1987). Regardless of the diversity of the jobs, participants clearly find (dis)similarities between jobs. As shown in Figure. 3a, participants do not randomly assign jobs to people, but rather, they cluster jobs into reasonable categories, and use the generalized category to guide decisions. For example, once participants discover Rekus are good custodians, they then assign Rekus to be cashiers and dishwashers even though they never have direct experience of Reku cashiers or Reku dishwashers because they perceive custodians as similar to cashiers and dishwashers. These two results highlight the unique contribution of this work, which is how feature-based exploration enables the emergence of multidimensional stereotypes. In sum, human behavioral data replicate the model predictions, showing that a stratified society emerges from participants

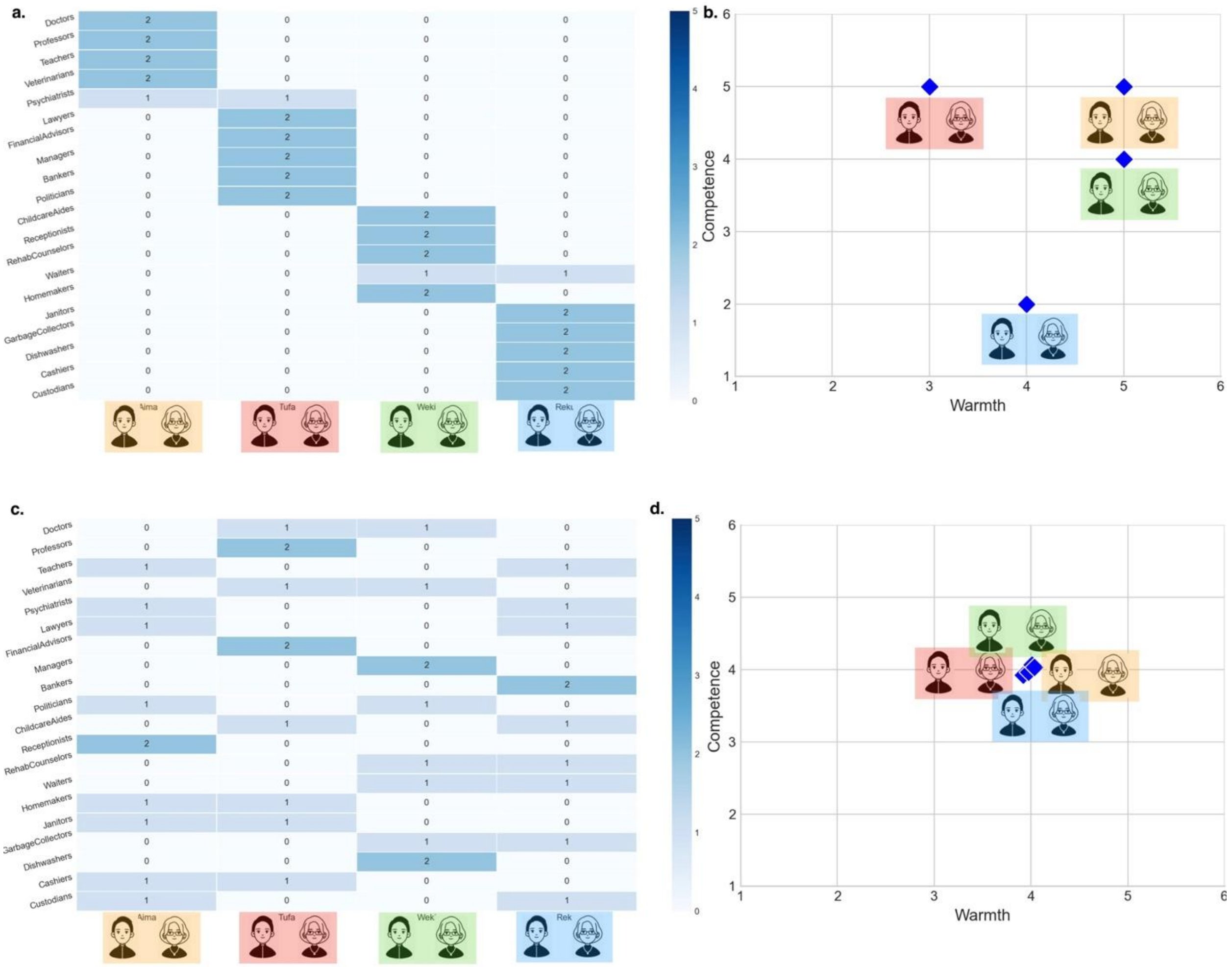


Figure. 3. Prototypes of stratified vs. diversified hiring choices and dissimilar vs. similar stereotypes. Panels a and b show results from participant #153, who was assigned to the adaptive exploration condition. This participant predominantly selects Aimas to work in high-status high-trust jobs, Tufas in high-status low-trust jobs, Wekis in low-status high-trust jobs, and Rekus in low-status low-trust jobs (a). As a result of such stratified choices, this participant thinks Aimas are warm (trustworthy) and competent, Tufas are competent but not warm, Wekis are incompetent but warm, and Rekus are neither competent nor warm (b). Panels c and d show results from participant #281, who was assigned to the random exploration condition. This participant observes the mayor selecting randomly (c). As a result, this participant thinks Aimas, Tufas, Wekis, and Rekus are similarly warm and competent (d).

acting adaptively to solve the explore-exploit tradeoff, and that this stratification leads to multidimensional stereotypes with a similar structure to those observed for real social groups.

Evaluating Interventions. If the implicit cost of exploration is the key mechanism that results in multidimensional stratification, intervening on this cost should reduce stratification and stereotypes. We studied three interventions to test this prediction: adding an exploration bonus, decreasing the reward probability, and imposing a random holdout. Each intervention addresses the implicit cost of exploration in a different way. First, adding a bonus to untried options directly incentivizes exploration (Bellemare et al., 2016). Second, decreasing the reward probability to make all groups less likely to yield rewards should make it less likely that people quickly encounter a successful group, meaning that they need to explore more. Third, randomly holding out some groups to make them unavailable forces exploration, making the cost of exploration irrelevant.

We initially tested these interventions using our computational model, which showed that all three interventions resulted in more diverse choices and more similarity among the estimated parameters of the groups (see *SI* for modeling results). We then tested these interventions in a behavioral experiment. Human participants were randomly assigned to one of the four conditions ($N = 807$): The control condition proceeds with the same hiring scenario as the adaptive exploration condition of our original experiment; the exploration bonus condition adds a diversity bonus, and it displays the sum of rewards from hiring decisions throughout the experiment; the lower reward condition decreases the underlying reward probabilities without an explicit change in instructions; the random holdout condition adds a travel restriction that randomly affects different groups, making two groups unclickable most of the time (see *SI* for experimental designs).

Consistent with the model, participants made more exploratory hiring when they were assigned to the exploration bonus ($b = .390$, 95% *CI* [.340, .440], $p < .001$), lower reward ($b = .402$, 95% *CI* [.355, .449], $p < .001$), and random holdout ($b = .319$, 95% *CI* [.272, .366], p

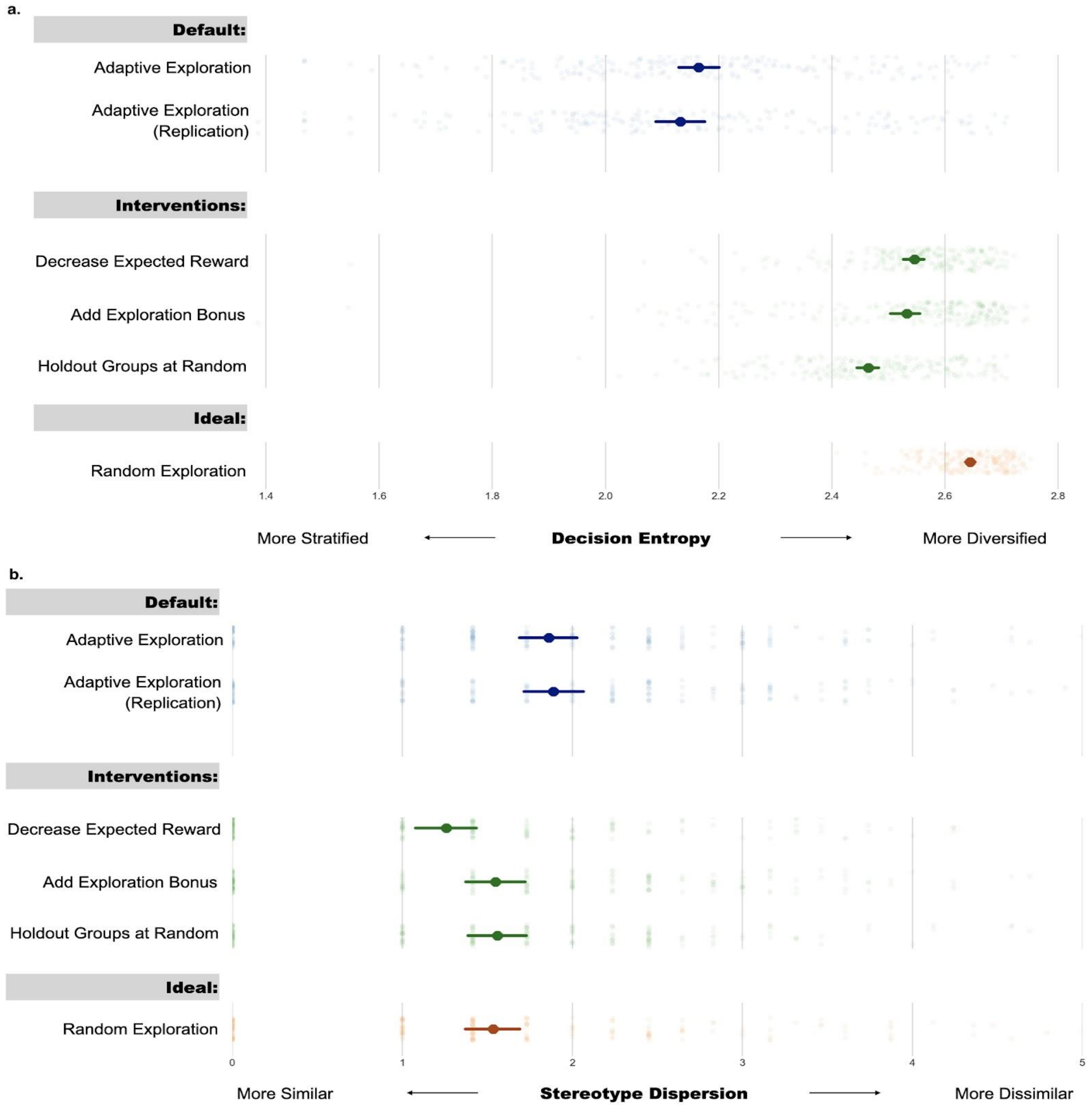


Figure. 4. Average treatment effects in human behavioral experiments. The vertical axis represents experimental conditions: The default panel with blue bars shows the adaptive exploration condition where participants make their own hiring decisions in the main study and replication in the mechanism study. The ideal panel with orange bars shows the random exploration condition where participants observe the mayor making random decisions. The intervention panel with green bars shows three interventions that manipulate the exploration cost to diversify choices and reduce stereotypes. The horizontal axis represents the average treatment effects for hiring choices in terms of choice entropy in panel a and stereotype dispersion in panel b. (a) shows more stratified choices to more diversified choices in the order of the default exploration, the interventions, and the random ideal condition. (b) shows more dissimilar to similar stereotypes in the order of the default exploration, the interventions, and the random ideal condition. In all graphs, error bars represent bootstrapped 95% confidence intervals.

< .001) conditions than those in the control condition (Figure. 4a. "Interventions" and "Default replication"). There are consistent, although weaker, treatment effects on the distances between the estimated parameters of the four Toma groups. Compared to the baseline, participants reveal smaller distances on the trustworthiness-competence space in the exploration bonus (b = -.339, 95% *CI* [-.603, -.074], p = .012), lower reward (b = -.693, 95% *CI* [-.959, -.427], p < .001), and random holdout (b = -.294, 95% *CI* [-.557, -.30], p = .029; Fig. 4b. "Interventions" and "Default replication") conditions. This pattern is robust for future hires and status-cooperation dimensions (see *SI* for more results). Interventions that change the cost of exploration are thus promising avenues for mitigating stratification and stereotypes.

Discussion

The mechanism of feature-based exploration we have introduced in this paper makes several innovative contributions. First, it provides a plausible explanation for the emergence of multidimensional stereotypes rather than those based purely on valence. Without assuming deficits in either decision-makers or the environmental samples, feature-based exploration explains how multidimensional stratification and stereotypes can emerge when decision-makers need to minimize exploration cost by both exploiting past experiences and generalizing from limited experiences to similar but not identical contexts. In an incentivized hiring experiment, using as many as 20 diverse real jobs, this mechanism is sufficient to reproduce the warmth-by-competence space that people use to represent real social groups. Learning that one group is good at doing one category of jobs and using that experience to guide category-sensitive decisions is adaptive to the individual because it minimizes exploration costs. Nonetheless, this strategy brings collateral damage to society - because it leaves other groups under-explored for certain types of jobs, resulting in stratification along dimensions that guide interpersonal interactions. Second, our intervention studies are the first to show that exploration cost *per se* is the operative variable. Introducing bonus rewards for diverse hires, assessing candidates using challenging tasks, and randomly making some groups unavailable for selection effectively reduces the cost of exploration, diversifies decisions, and reduces stereotypes.

Our proposed mechanism complements but differs from prior theories on the origin of stereotypes, as follows. (a) The motivation to maximize self-interest can be orthogonal to the motivation to maintain group identity or hierarchy (e.g., Brewer, 1999). Identifying the causes of stratification and stereotypes as pursuing self-interest with exploration yields very different interventions. Complementing strategies such as creating a common ingroup identity (Gaertner & Dovidio, 2009), our proposal suggests changes in the reward structure for exploration. Consistent with the call for structural changes to redress social bias, our mechanism provides concrete ideas such as introducing bonus rewards for diverse hires. (b) A lack of exploration differs from confirmation bias or metacognitive myopia (e.g., Hamilton & Gifford, 1976). To see why, disentangle two different goals. The incentive in our task is to maximize rewards (earn as many points as possible), whereas the incentive in confirmation bias and metacognitive myopia is to strengthen beliefs (learn the underlying principles as accurately as possible). Although it has been assumed that to maximize rewards one needs to maximize accuracy, we show that the two goals do not always align. Hence, inaccuracy can arise not as a cognitive limitation, but as a side-effect of trying to maximize rewards (see also Le Mens & Denrell, 2011; Rich & Gureckis, 2018). (c) Our proposed mechanism does not depend on asymmetric population sizes when one group is more accessible than other groups (e.g., Alves et al., 2018). Adding unbalanced population size may exacerbate this effect; however, one should not forget that the definitions of majority and minority are not fixed either. Rather than starting with a fixed majority/minority representation, our mechanism provides a process that may create such asymmetry: Individuals who are not explored enough become the numerical minority. (d) Our proposed mechanism does not endorse stereotype accuracy at all (e.g., Jussim, 2017), because we showed inaccurate stereotypes emerge even when the ground truth is otherwise.

Most importantly, none of the above theories demonstrably explains why stereotypes have more than one dimension. In contrast, we find the diverse contents of stereotypes associated with social groups could be a result of generalization based on socially constructed features of different jobs. Given that identical situations are rarely encountered twice, the ability to generalize is a crucial adaptive mechanism for humans (Shepard, 1987; Schulz et al., 2018). However, when this generalization process is coupled with decisions to balance exploration and exploitation, it can lead to wrongful association of certain features with specific groups. Absent evidence from less-explored alternatives, people might consistently apply these generalized features in future judgments, laying the ground for multidimensional stereotypes. If jobs or social roles were restricted to a single valence dimension, we would expect to see stereotypes represented merely by positivity and negativity. Yet, our empirical evidence – a large sample of ecologically valid jobs – indicates that human participants perceive jobs varying across at least two dimensions, supporting the plausibility of multidimensional stereotype framework.

Social scientists have studied diversity and stereotypes from either an individual or a structural lens. However, the new mechanism we have identified suggests that the culprit may be an interaction of the two. It challenges the common assumption that unjust systems are either the result of prejudiced or cognitively stressed decision-makers, or the result of power-maintaining or undiversified organizational arrangements. Instead, it highlights the possibility that unjust systems can also be created by locally adaptive, reward-maximizing decision-makers. A company merely pursuing its profit can hire certain groups of workers for specialized tasks but under-explore other groups for inexperienced tasks (Li et al., 2020). A university merely pursuing a higher ranking for research can admit certain kinds of researchers for particular disciplines but under-explore other combinations (Wapman et al., 2022). These reasonable local decisions in the short term can create stratified broader societal structures in the long term.

Some real-world policy implications of this idea range well beyond employment discrimination. For example, one pertains to refugee resettlement. Policymakers and social scientists, leveraging large-scale datasets and machine-learning algorithms, propose allocating refugees with similar demographic features to specific locations for similar jobs based on past success (Bansak et al., 2016). Such a plan can be suitable for refugees in the short term because it brings more satisfaction and contributes to the local economy. However, this plan, our model predicts, will cause future damage in the form of multidimensional stereotyping and data-driven discrimination.

The exploration-cost mechanism that produces stereotypes in humans also provides a psychological analog of fairness concerns in artificial intelligence. For instance, recommendation algorithms often attempt to infer user preferences based on their past behaviors. However, these algorithms may inadvertently limit exposure to diverse options, making some unreachable to users (Dean et al., 2020). While optimizing customer engagement may be an adaptive strategy for the local algorithm, it simultaneously perpetuates stratification in the global online system.

Stereotypes are shared cultural beliefs, and segregation is a collective endeavor. Future work should study how idiosyncratic and biased individual experiences become entrenched, not mitigated, within collective systems (Martin et al., 2014; Lyons & Kashima, 2003). Our approach extracts the minimal conditions under which stereotypes can emerge, but it needs real-world corroboration. Future work can use historical, immigration, or organizational datasets to examine adaptive exploration in everyday choices (Card et al., 2022; Charlesworth et al., 2022). Costly exploration should be added to the list of psychological mechanisms that can lead to stereotypes, creating an opportunity for future research that integrates these different mechanisms (Almaatouq et al., 2022). However, continuing to ignore the role of exploration in the creation of stereotypes will reinforce the very injustices that we seek to eradicate. Scientists and practitioners should design systems that facilitate exploration in social decision-making, and the interventions explored here provide a first step in that direction.

Materials and Methods

Experimental details, dataset construction, analysis details, formal modeling, and computational simulations are provided in the Supplementary Information.

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.

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2022). Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences. *Behavioral and Brain Sciences*, 1-55.

Alves, H., Koch, A., & Unkelbach, C. (2018). A cognitive-ecological explanation of intergroup biases. *Psychological Science*, *29*(7), 1126-1133.

Altemeyer, B. (1983). *Right-wing authoritarianism*. Univ. of Manitoba Press.

Bai, X., Fiske, S. T., & Griffiths, T. L. (2022). Globally inaccurate stereotypes can result from locally adaptive exploration. *Psychological Science*, *33*(5), 671-684.

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.

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, *359*(6373), 325-329.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, *29*.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391.

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

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., ... & Jurafsky, D. (2022). Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, *119*(31), e2120510119.

Charlesworth, T. E., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from Google Books. *Proceedings of the National Academy of Sciences*, *119*(28), e2121798119.

Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, *24*.

Dean, S., Rich, S., & Recht, B. (2020, January). Recommendations and user agency: the reachability of collaboratively-filtered information. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 436-445).

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

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*, *46*(4), 735.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological review*, *107*(4), 659.

Fiske, S. T., Cuddy, A. J., 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.

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, *111*(supplement_4), 13593-13597.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. McGraw-Hill.

Gaertner, S. L., & Dovidio, J. F. (2009). A common ingroup identity: A categorization-based approach for reducing intergroup bias. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 489–505). Psychology Press.

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.

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.

Jussim, L. (2017). Précis of Social Perception and Social Reality: Why accuracy dominates bias and self-fulfilling prophecy. *Behavioral and Brain Sciences*, *40*, e1.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: observations of groups' roles shape stereotypes. *Journal of personality and social psychology*, *107*(3), 371.

Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the Stereotype Content Model. *International Journal of Intercultural Relations, 30*(6), 751–768.

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), 379–392.

Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010, April). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web* (pp. 661-670).

Li, D., Raymond, L. R., & Bergman, P. (2020). *Hiring as exploration* (No. w27736). National Bureau of Economic Research.

Lyons, A., & Kashima, Y. (2003). How are stereotypes maintained through communication? The influence of stereotype sharedness. *Journal of personality and social psychology*, *85*(6), 989.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of personality and Social Psychology*, *66*(1), 37.

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.

McCauley, C. R., Jussim, L. J., & Lee, Y. T. (1995). *Stereotype accuracy: Toward appreciating group differences*. American Psychological Association.

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Journal of personality and social psychology*.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233-248.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, *67*(4), 741.

Rich, A. S., & Gureckis, T. M. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General*, *147*(11), 1553.

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.

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and stereotyping: Cognitive constraints on the construction of meaningful social impressions. *European review of social psychology*, *11*(1), 145-175.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317–1323.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623-1626.

Trope, Y., & Thompson, E. P. (1997). Looking for truth in all the wrong places? Asymmetric search of individuating information about stereotyped group members. *Journal of Personality and Social Psychology, 73*(2), 229–241.

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.

Wapman, K. H., Zhang, S., Clauset, A., & Larremore, D. B. (2022). Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature*, *610*(7930), 120-127.

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of personality and social psychology*, *112*(5), 696.

Supplementary Information for

Multidimensional Stereotypes Emerge Spontaneously When Exploration is Costly

Anonymized Authors

The PDF file includes:

Other Supplementary Materials for this manuscript include the following:

- Movies S1 to S6.
- Data S1 to S3.
- Code S1 to S4.

Note. To facilitate the understanding of our mathematic models of contextual multi-armed bandit and Bayesian inference, we made a tutorial video to explain the process as intuitively as possible. Interested readers can find it from the link for Movie S6.

Materials and Methods

Human Experiments.

In this section, we report additional details about human experiments. All studies are approved by the Institutional Review Board at [mask] University under protocol number 13065. All studies are preregistered at https://osf.io/6p8wu/registrations; author-identifiable information is included. Additional pilot studies for minor tweaks, such as pilot experiments with selected prototypical jobs, stimulus choices with respect to wording, and various intervention prompts are not included in this report but are documented on the preregistration site. Corresponding to the main text, Study S1 reports a systematic analysis of stimulus jobs, Study S2 reports the main hiring experiment, and Study S3 reports mechanism experiments.

Study S1. Stimuli: Jobs and Dimensions.

Participants. We recruited $N = 100$ online workers from the Cloud Research high-quality subject pool who speak English as their first language and are older than 18 years old. This sample size was calculated based on prior work in warmth and competence research (*3, 6, 23*), The average age was 41; 50% female, 50% male; 85% White, 7% Black, 4% Asian, and 71% participants hold some college or bachelor's degree, reflecting typical demographic characteristics of online American workers for psychological studies.

Materials. In the survey, we asked participants to rate 24 occupations in terms of their perceived status, cooperation, competence, and warmth. The 24 jobs were selected based on the US Bureau of Labor Statistics Occupation Outlook Handbook published in 2020, social perceptions about common jobs (*29*), and common beliefs about occupational roles (*6*). According to prior work, we anticipated the following categories: perceived warm and competent jobs include Doctors, Veterinarians, Professors, Teachers, Psychiatrists, and Computer Scientists; perceived cold but competent jobs include Lawyers, Managers, Financial Advisors, Bankers, Politicians, and Fashion Designers; perceived warm but incompetent jobs include Childcare Aides, Receptionists, Rehabilitation Counselors, Waiters, Homemakers, and Nursing Assistants; and perceived cold and incompetent jobs include Janitors, Custodians, Truck Drivers, Garbage Collectors, Dishwashers, and Cashiers.

For each job, we asked participants to rate the following eight items from 1 (not at all) to 5 (extremely), as viewed by American society (not their personal beliefs; *7*). Status items: How economically successful/well-educated have members of this occupation been? Cooperation items: If resources go to members of this occupation, to what extent does that take resources away from the rest of society/How much does special treatment given to members of this occupation make things more difficult for other groups in society? Competence items: How capable/confident are members of this occupation? Warmth items: How friendly/trustworthy are members of this occupation? Movie S1 shows participant experiences during the task.

Results. K-means clustering is an unsupervised algorithm that was used to partition the 24 jobs into 4 clusters in which each job belongs to the cluster with the nearest cluster centroid (*23*). We used Lloyd's algorithm to iteratively refine the cluster assignments. The algorithm proceeds by alternating between the two steps of assignment and update. During the assignment step, it assigns each observation to the cluster with the nearest mean or the least squared Euclidean distance. During the update step, it recalculates the means for observations assigned to each cluster. The algorithm converges when the assignments no longer change, giving the final cluster assignments as the output. Data in Data S1, Analysis code in Code S1, and results in Fig. S1.

To reduce noise, we removed four ambiguous jobs: computer scientists could belong to high-warmth high-competence or low-warmth high-competence clusters, fashion designers could belong to high-warmth high-competence or low-warmth high competence, nursing assistants could be high-warmth high-competence or high-warmth low-competence, and truck drivers could be low-warmth low-competence or low-warmth high-competence. Hence, in the main experiment, we used the refined 20 jobs as the experimental stimuli. This use of real-world jobs and human judgments improves ecological validity.

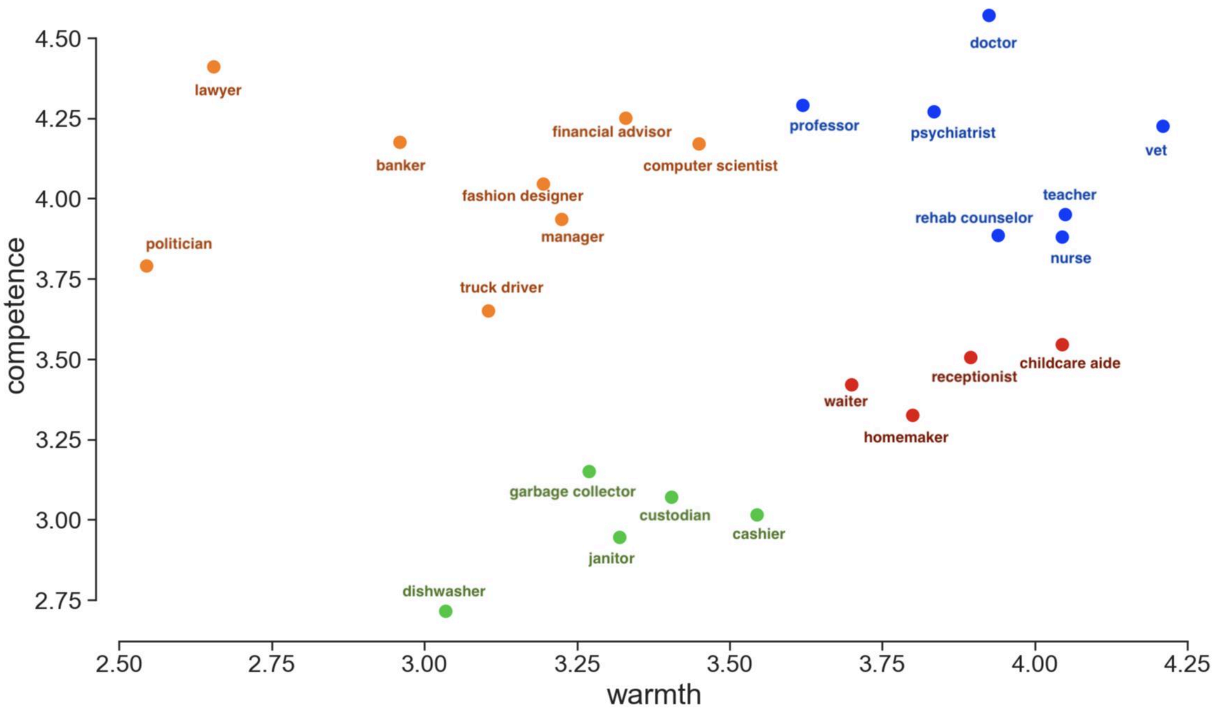


Fig. S1. Emerging clusters of common jobs in American society along dimensions of status or competence, and cooperation or warmth. Values on the axis reflect estimated values, on a scale from 1 to 5, of each job along the two dimensions. Colors indicate cluster assignments as calculated via the KMeans clustering algorithm.

Study S2. Main Experiment: Hiring Consultant for Toma City

Participants. We recruited *N* = 403 online workers from the Cloud Research high-quality subject pool with the same selection criteria as in Study S1. This sample size was calculated based on a pilot study (see details in this anonymized preregistration). The average age was 40; 51% female, 46% male, and 1% non-binary; 74% White, 10% Black, 6% Hispanic, 5% Asian, and 4% multi-racial; 75% of participants hold some college or bachelor’s degree; the average political orientation was slightly liberal with an average score of 3.94 on a scale from 1 extremely conservative to 6 extremely liberal.

Materials. This experiment extends the context-free multi-armed bandit behavioral experiment in (*27*) to test the emergence of multi-dimensional stratification using hiring decisions. In the cover story, participants learn that they will play a game with made-up people from a made-up city. Toma City has around 100,000 residents; they come from four ancestral villages: Tufa, Aima, Reku, and Weki. Participants are hired as a consultant by the mayor of Toma City, and their task is to recommend Toma people for various jobs, out of 20 jobs, in 40 sequential decisions. After each recommendation, participants will learn whether it is a good choice or not. A perfect fit earns 1 point whereas a bad fit earns 0 points. The more points the participants earn, the more bonus they get (1 point = 1 cent), in addition to their base pay ($3 for a 20-minute task). In the game phase, participants see "Job Opening: Doctors" in the first round. They then must select one member from Tufa, Aima, Reku, and Weki groups. On the next page, they see either "You earned 1 point" or "You earned 0 points." Participants then proceed to the second round, recommend another randomly generated job, and receive feedback. There are 40 trials in total, and after finishing all decisions, participants are asked to answer some questions. First, they are asked one generalization question: "Imagine there are 100 new individuals from each village group applying for the jobs. Enter how many of them you would recommend for each job." They enter values for the four groups for the twenty jobs. Next, participants are asked about their impressions of the four groups, on a scale from 1 (not at all) to 5 (extremely): "Tufas/Aimas/Rekus/Wekis, in general, seem to be economically successful/interested in helping others/competent or confident/friendly or trustworthy."

As straightforward as the experiment appears, we made four critical decisions with the goal of minimizing other psychological mechanisms in crafting this experiment. First, we minimized group-serving motivations such as ingroup favoritism (e.g., *8*) or social dominance (e.g., *10*) by assigning no prior group membership to any of our participants. In the spirit of the minimal group paradigm, the use of novel groups achieved this goal. Second, we tried to minimize the cognitive load (e.g., *13*) by reducing the number of trials in this study, visual representations in addition to abstract group names, and the overall presentation of the hiring interface. Third, to rule out population size as one alternative explanation (e.g., *17*), in the backend, we prepared all groups with equal population sizes, that is 40 Tufas, 40 Aimas, 40 Rekus, and 40 Wekis will be available if selected. Fourth, just as in the model simulation, we set the true success probability for the four groups in Toma City for the twenty jobs as high and identical, with a 90% success rate for all job-group combinations. This manipulation eliminated the alternative explanation of ground truth differences (e.g., *21*). The average completion time is 18 minutes. Participants in general enjoyed this task as many left comments saying they had never done a task like this before and it made them think. Data in Data S2 and Movie S2 adaptive and random show participant experiences during the task.

Treatment. The key treatment is the method of exploration. There are two conditions in this hiring experiment. In the experimental condition, participants made hiring decisions as they wished in the infrastructure described above. In the control condition, participants did not have the opportunity to make their own decisions. Instead, they learned that "The mayor will make one recommendation each time, and you can observe the mayor's decision." From the backend, the game infrastructure selected each group randomly at each time, to mimic the experience of random-decision. After 40 trials of hiring decisions, participants in both conditions continued to make future hires and provided impressions about the groups as described above.

Results. We estimated OLS regressions in which we regressed our outcomes – choice entropy during the 40-trial game, choice entropy of future hires, dispersion of estimated status and cooperation, and dispersion of estimated competence and warmth – over our treatment indicator (β), controlling for respondents' age, gender, race, education, and political orientation. Our main quantity of interest is on identifying β representing the average treatment effect of the exploration strategy on participants' hire decisions and impressions about Toma groups. Results summary in Table S1. Analysis code in Code S2.

Table S1. Average Treatment Effects.

	β	***t***	***p* > | *t* |**	**[.025, .975]**
Choice entropy: 40-trial current hires				
Intercept (=Adaptive)	2.165	162.222	.000	[2.138, 2.191]
Random	0.480	25.360	.000	[0.443, 0.518]
Choice entropy: 40-trial current hires				
Intercept (=Adaptive, Female, Black)	2.205	33.446	.000	[2.076, 2.335]
Random	0.476	24.311	.000	[0.437, 0.514]
Gender (=Male)	-0.017	-0.880	0.379	[-0.056, 0.021]
Gender (=Nonbinary)	0.097	1.006	0.315	[-0.093, 0.288]
Race (=Asian)	-0.079	-1.448	0.148	[-0.187, 0.028]
Race (=Caucasian)	-0.084	-2.559	0.011	[-0.148, -0.019]
Race (=Hispanic)	-0.042	-0.815	0.416	[-0.142, 0.059]
Race (=Multiracial)	-0.009	-0.152	0.879	[-0.125, 0.107]
Age	-0.000	-0.109	0.913	[-0.002, 0.002]
Education	0.004	0.370	0.711	[-0.017, 0.024]
Political Orientation	0.007	1.017	0.310	[-0.007, 0.021]
Choice entropy: 400 future hires				
Intercept (=Adaptive)	2.169	89.752	.000	[2.122, 2.217]
Random	0.438	12.754	.000	[0.370, 0.505]
Choice entropy: 400 future hires				
Intercept (=Adaptive, Female, Black)	2.331	19.363	.000	[2.094, 2.568]
Random	0.432	12.104	.000	[0.362, 0.503]

Gender (=Male)	0.014	0.389	0.698	[-0.057, 0.085]
Gender (=Nonbinary)	0.240	1.359	0.175	[-0.107, 0.588]
Race (=Asian)	-0.152	-1.521	0.129	[-0.349, 0.045]
Race (=Caucasian)	-0.132	-2.203	0.028	[-0.249, -0.014]
Race (=Hispanic)	-0.168	-1.796	0.073	[-0.351, 0.016]
Race (=Multiracial)	-0.064	-0.600	0.549	[-0.275, 0.147]
Age	-0.002	-1.128	0.260	[-0.005, 0.001]
Education	-0.007	-0.375	0.708	[-0.045, 0.030]
Political Orientation	0.011	0.850	0.396	[-0.014, 0.036]

Stereotype dispersion: cooperation and status

Intercept (=Adaptive)	2.637	27.428	.000	[2.448, 2.826]
Random	-0.747	-5.476	.000	[-1.016, -0.479]

Stereotype dispersion: cooperation and status

Intercept (=Adaptive, Female, Black)	3.360	6.978	.000	[2.413, 4.307]
Random	-0.753	-5.271	.000	[-1.034, -0.472]
Gender (=Male)	0.037	0.253	0.801	[-0.247, 0.320]
Gender (=Nonbinary)	-0.116	-0.164	0.869	[-1.507, 1.275]
Race (=Asian)	-0.047	-0.117	0.907	[-0.833, 0.740]
Race (=Caucasian)	0.126	0.527	0.598	[-0.344, 0.596]
Race (=Hispanic)	0.392	1.050	0.294	[-0.342, 1.126]
Race (=Multiracial)	0.087	0.202	0.840	[-0.757, 0.931]
Age	-0.004	-0.721	0.472	[-0.016, 0.008]

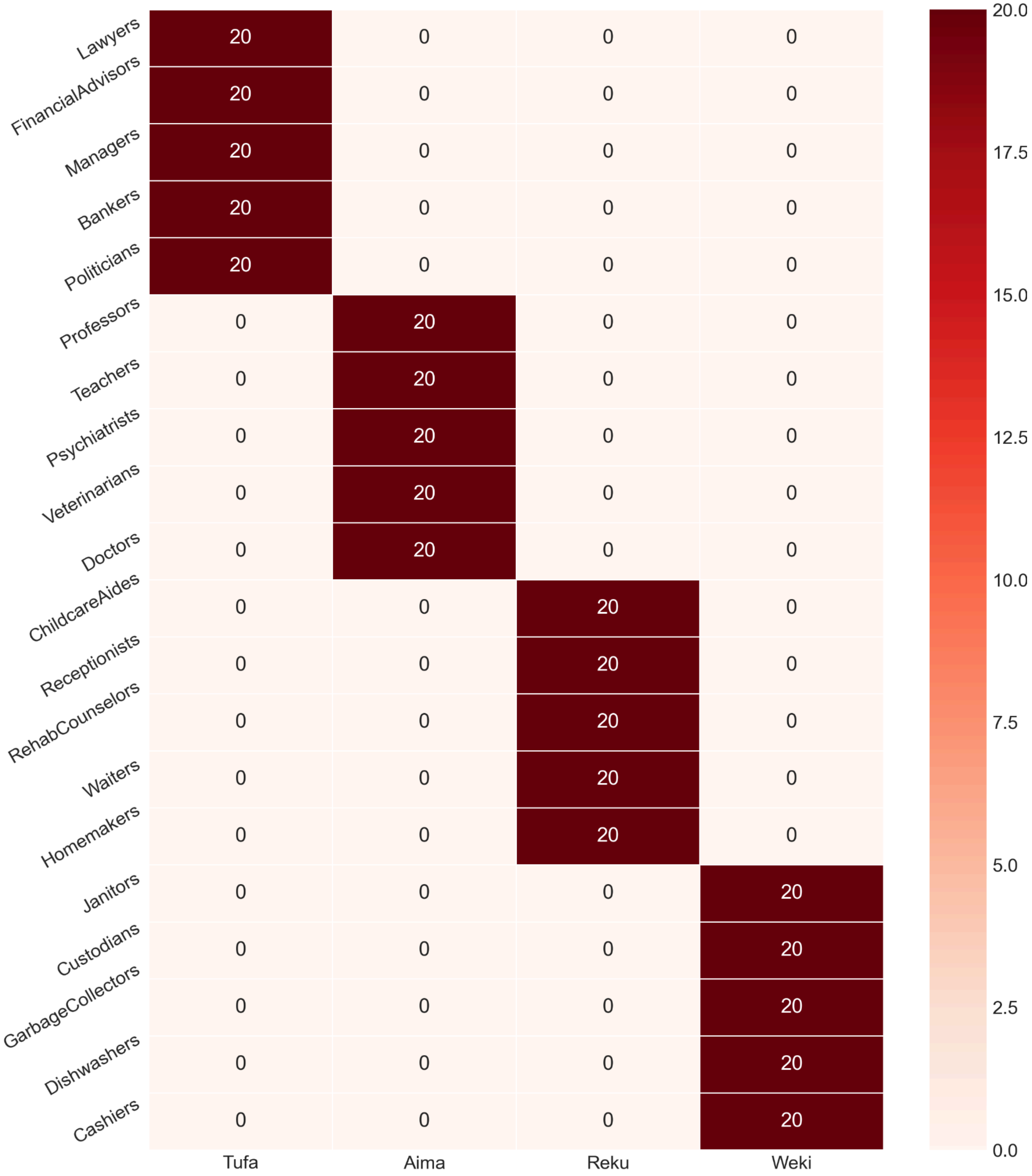
Education	-0.127	-1.660	0.098	[-0.277, 0.023]
Political Orientation	-0.056	-1.096	0.274	[-0.157, 0.045]

Stereotype dispersion: warmth and competence				
Intercept (=Adaptive)	1.861	21.520	.000	[1.691, 2.031]
Random	-0.327	-2.665	.008	[-0.568, -0.086]

Stereotype dispersion: warmth and competence				
Intercept (=Adaptive, Female, Black)	2.397	5.505	.000	[1.541, 3.254]
Random	-0.343	-2.653	.008	[-0.597, -0.089]
Gender (=Male)	-0.002	-0.018	0.986	[-0.259, 0.255]
Gender (=Nonbinary)	-0.201	-0.314	0.753	[-1.459, 1.057]
Race (=Asian)	-0.204	-0.564	0.573	[-0.916, 0.507]
Race (=Caucasian)	0.050	0.233	0.816	[-0.375, 0.475]
Race (=Hispanic)	-0.013	-0.039	0.969	[-0.677, 0.651]
Race (=Multiracial)	0.023	0.059	0.953	[-0.741, 0.786]
Age	-0.007	-1.340	0.181	[-0.018, 0.003]
Education	-0.056	-0.814	0.416	[-0.192, 0.080]
Political Orientation	-0.013	-0.269	0.788	[-0.104, 0.079]

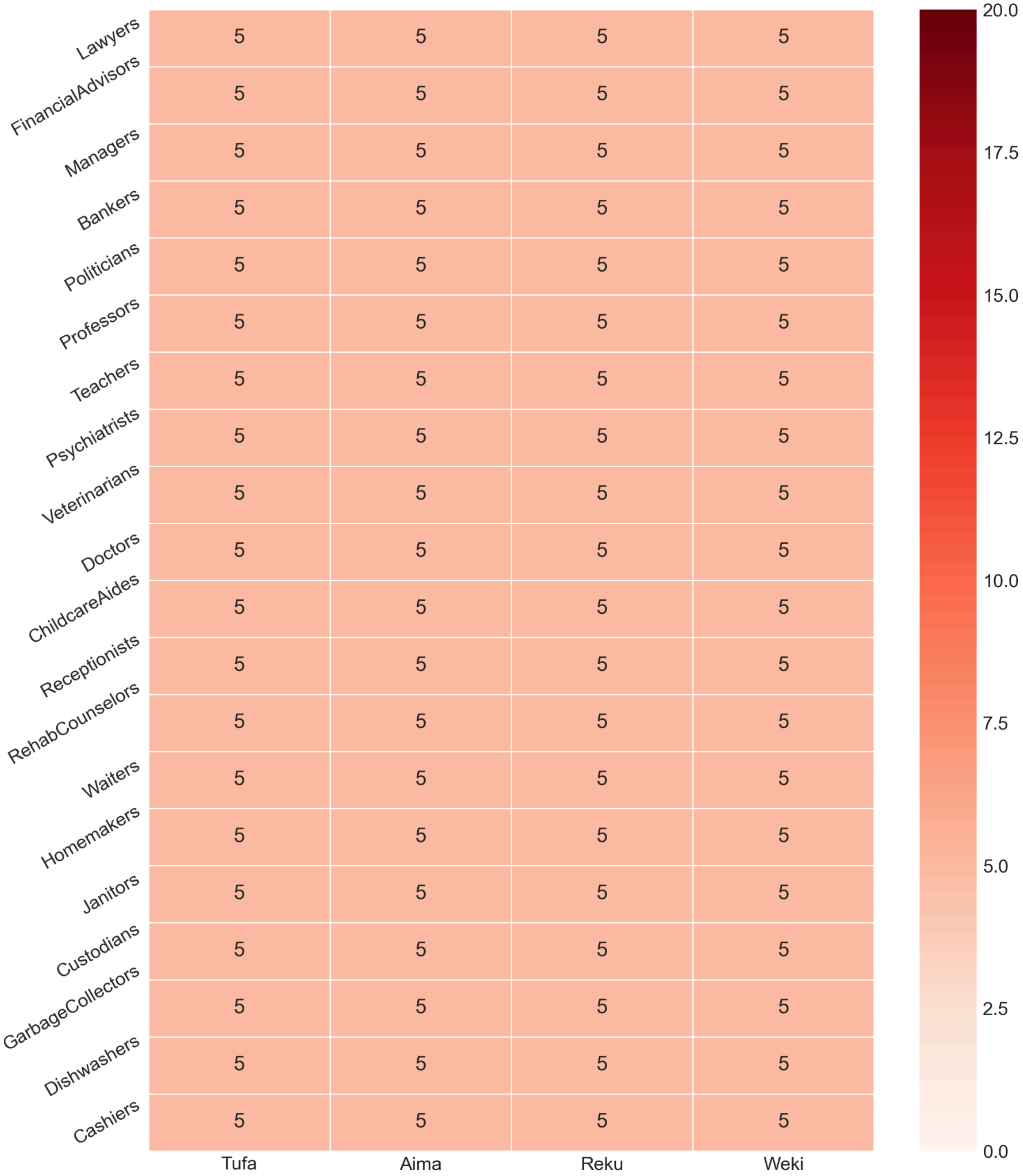
Note. Estimates are based on an OLS model without (first panels) and with (second panels) covariate variables of age, gender, race, education, and political orientation. $N = 403$.

We plotted exemplar participants from the 40-trial condition in the main text, here we provide their hiring decisions in future jobs and along dimensions of status and cooperation (Figs. S2 - S5). For complete participants, see Data S2 and use Code S2.



	Tufa	Aima	Reku	Weki
Lawyers	20	0	0	0
FinancialAdvisors	20	0	0	0
Managers	20	0	0	0
Bankers	20	0	0	0
Politicians	20	0	0	0
Professors	0	20	0	0
Teachers	0	20	0	0
Psychiatrists	0	20	0	0
Veterinarians	0	20	0	0
Doctors	0	20	0	0
ChildcareAides	0	0	20	0
Receptionists	0	0	20	0
RehabCounselors	0	0	20	0
Waiters	0	0	20	0
Homemakers	0	0	20	0
Janitors	0	0	0	20
Custodians	0	0	0	20
GarbageCollectors	0	0	0	20
Dishwashers	0	0	0	20
Cashiers	0	0	0	20

Fig. S2. Illustrative participants *ID* = 153 in the adaptive exploration condition made more stratified and less diverse future hiring decisions.



	Tufa	Aima	Reku	Weki
Lawyers	5	5	5	5
FinancialAdvisors	5	5	5	5
Managers	5	5	5	5
Bankers	5	5	5	5
Politicians	5	5	5	5
Professors	5	5	5	5
Teachers	5	5	5	5
Psychiatrists	5	5	5	5
Veterinarians	5	5	5	5
Doctors	5	5	5	5
ChildcareAides	5	5	5	5
Receptionists	5	5	5	5
RehabCounselors	5	5	5	5
Waiters	5	5	5	5
Homemakers	5	5	5	5
Janitors	5	5	5	5
Custodians	5	5	5	5
GarbageCollectors	5	5	5	5
Dishwashers	5	5	5	5
Cashiers	5	5	5	5

Fig. S3. Illustrative participants *ID* = 281 in the random exploration condition made less stratified and more equal future hiring decisions.

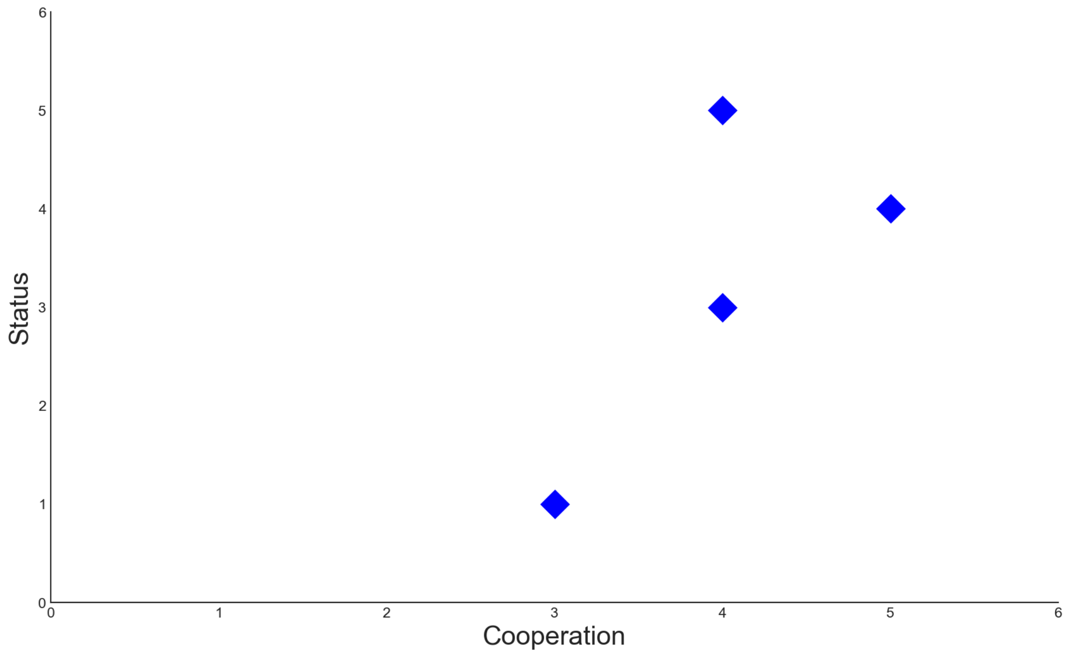


Fig. S4. Illustrative participants *ID* = 153 in the adaptive exploration condition showed more dispersed mental maps along the social status and cooperative intent dimensions.

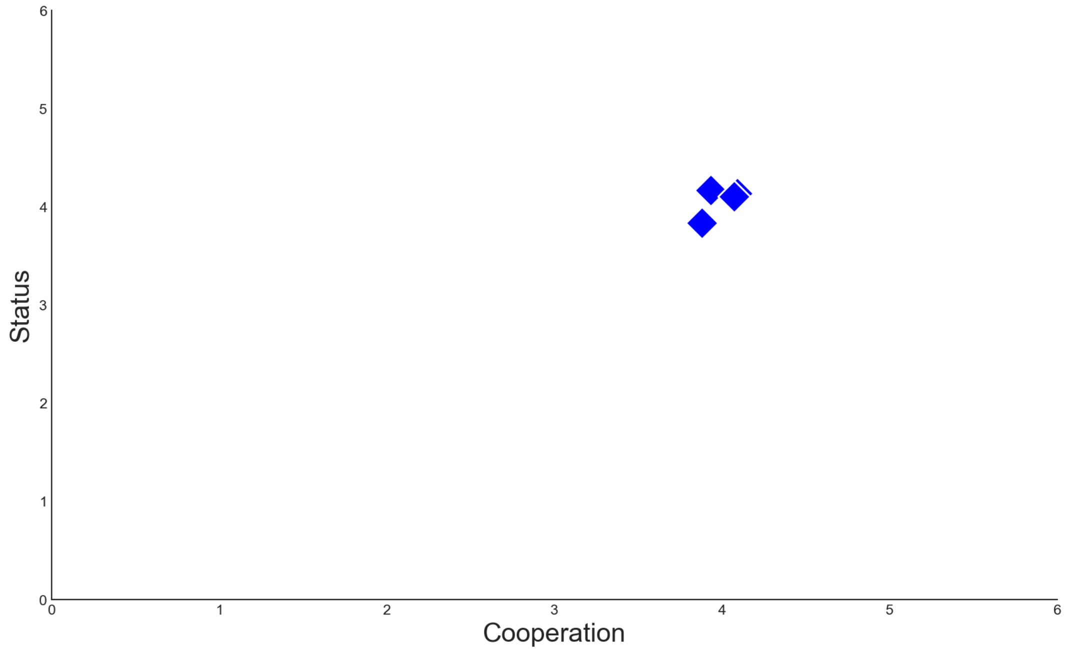


Fig. S5. Illustrative participants *ID* = 281 in the random exploration condition showed less dispersed mental maps along the social status and cooperative intent dimensions.

Study S3. Mechanism Experiment: Interventions for Exploration

Participants. We recruited $N = 807$ online workers from Connect, a new platform for paid online studies hosted by Cloud Research. The reason to switch from Cloud Research to Connect is rather practical as Cloud Research only allows large-scale payment if we recruit MTurk participants, otherwise we have a daily payment limitation. We used the same selection criteria as in previous experiments, while also imposing gender balance given it is a convenient function on Connect. The sample size was again to ensure 200 participants per condition, to be consistent with Study 2 (see details in this anonymized preregistration). The average age was 40; 50% female, 50% male; 67% White, 11% Black, 9% Asian, 6% Hispanic, and 4% multi-racial; 71% of participants hold some college or bachelor's degree; the average political orientation was slightly liberal with an average score of 3.98 on a scale from 1 extremely conservative to 6 extremely liberal. The average score for this task was 4.7 out of 5, indicating acceptable engagement among participants. In addition to the main mechanism, we also piloted similar experiments to nail down the wording for the interventions; details can be found in pre-registration reports titled "mechanism pilot."

Materials. The experiment materials were largely identical to the main experiment described above. All data from this experiment are in Data S3. The baseline condition is identical, whereas the intervention conditions contain different designs mainly in the cover story, the hiring decisions, and the feedback pages, as follows.

In the exploration bonus condition, after participants read general instructions about the city, jobs, their roles, and points, and before they started the game, they saw a new page titled "Diversity Bonus." They read: "Recently Toma City launched a hiring initiative. The mayor will pay an extra bonus for more variety in who you hire. The bonus decreases for each hire of a person from a group that has previously been hired for that job. Your total earnings will be the sum of rewards from making suitable hires and the diversity bonus." After they made one hiring decision, they received feedback as participants in the baseline condition. However, the bonus was now calculated as the sum of their actual reward (1 or 0) and a diversity bonus of $1/(N+1)$, N being the times of the same group being recommended for the same cluster of jobs. Therefore, rather than displaying an integer value of 1 or 0, this page showed floating numbers such as 2 (if the selection is completely new, 1/(0+1), and the selection is good, 1), 1.5 (if the selection is good, but it is the second time being recommended, 1/(1+1)), and so on. Movie S3 shows participants' experiences in this condition.

In the reward rate condition, the instructions did not change at all. The only difference was the underlying expected reward which changed from 90% to 10%. For example, among 40 available members from village Tufa, a 90% success rate means 36 of them would return a reward of 1 for the jobs being recommended, versus a 10% success rate means only 4 of them would be successful. Note that participants did not have access to this information before the game, and the only way to figure this out was through experience. Movie S4 shows participants' experiences in this condition.

In the random holdout condition, after participants read general instructions and before they started the game, they saw a new page titled "Travel Restrictions." They read: "Due to recent travel restrictions, not all villagers are able to come to work at all times. Sometimes your selected members might become unavailable; if so, you need to choose from the available members." To reflect this change, on their hiring page, 90% of all trials made 2 out of 4 groups not clickable indicating those two groups are not available due to travel restrictions. It was a random selection of which two groups to disable and on which trials participants encountered travel restrictions. Movie S5 shows participants' experiences in this condition.

Results. We used the same analysis strategy as in Study 2 in which we estimated OLS regressions by regressing our outcomes - choice entropy during the 40-trial game, choice entropy of future hires, dispersion of estimated status and cooperation, and dispersion of estimated competence and warmth - over our treatment indicator (β), controlling for respondents' age, gender, race, education, and political orientation. Our main quantity of interest is on identifying β representing the average treatment effect of the baseline default hiring and each of the three proposed interventions - exploration bonus, reward rate, random holdout - on participants' hire decisions and impressions about Toma groups. Given that we test for three hypotheses, the analysis used Bonferroni correction of the alpha level at 0.01. More precisely, the threshold should be the original alpha value divided by the number of comparisons, that is 0.05/3 = 0.0167. Results summary in Tables S2 - S4. Analysis code in Code S3.

Table S2. Average Treatment Effects (Exploration Bonus)

	β	*t*	*p* > | *t* |	**[.025, .975]**
Choice entropy: 40-trial current hires				
Intercept (=Adaptive)	2.133	118.962	.000	[2.097, 2.168]
Exploration Bonus	0.400	16.168	.000	[0.352, 0.449]
Choice entropy: 40-trial current hires				
Intercept (=Adaptive, Female, Black)	2.156	27.396	.000	[2.001, 2.310]
Exploration Bonus	0.396	15.484	.000	[0.346, 0.447]
Gender (=Male)	-0.022	-0.820	0.413	[-0.073, 0.030]
Race (=Asian)	-0.114	-2.108	0.036	[-0.221, -0.008]
Race (=Caucasian)	-0.153	-3.410	0.001	[-0.241, -0.065]
Race (=Hispanic)	-0.076	-1.223	0.222	[-0.197, 0.046]
Race (=Multiracial)	-0.125	-1.778	0.076	[-0.263, 0.013]
Age	0.001	1.033	0.302	[-0.001, 0.003]
Education	0.008	0.943	0.346	[-0.009, 0.026]
Political Orientation	0.011	1.281	0.201	[-0.006, 0.028]

Choice entropy: 400-trial future hires				
Intercept (=Adaptive)	2.166	79.479	.000	[2.113, 2.220]
Exploration Bonus	0.368	9.768	.000	[0.294, 0.442]
Choice entropy: 400-trial future hires				

Intercept (=Adaptive, Female, Black)	1.955	16.418	.000	[1.721, 2.189]
Exploration Bonus	0.366	9.445	.000	[0.290, 0.442]
Gender (=Male)	-0.028	-0.710	0.478	[-0.106, 0.050]
Race (=Asian)	-0.078	-0.954	0.341	[-0.240, 0.083]
Race (=Caucasian)	-0.124	-1.824	0.069	[-0.258, 0.010]
Race (=Hispanic)	-0.109	-1.163	0.246	[-0.294, 0.075]
Race (=Multiracial)	0.081	0.756	0.450	[-0.129, 0.290]
Age	0.003	2.068	0.039	[0.000, 0.007]
Education	0.020	1.513	0.131	[-0.006, 0.047]
Political Orientation	0.029	2.272	0.024	[0.004, 0.054]

Stereotype dispersion: cooperation and status				
Intercept (=Adaptive)	2.618	24.191	.000	[2.405, 2.830]
Exploration Bonus	-0.774	-5.179	.000	[-1.068, -0.480]
Stereotype dispersion: cooperation and status				

Intercept (=Adaptive, Female, Black)	2.628	5.506	.000	[1.690, 3.567]
Exploration Bonus	-0.788	-5.074	.000	[-1.093, -0.482]
Gender (=Male)	0.104	0.657	0.512	[-0.208, 0.416]
Race (=Asian)	0.483	1.466	0.143	[-0.165, 1.132]
Race (=Caucasian)	0.039	0.142	0.887	[-0.497, 0.575]
Race (=Hispanic)	-0.325	-0.865	0.388	[-1.065, 0.414]
Race (=Multiracial)	0.254	0.427	0.552	[-0.585, 1.093]
Age	0.006	0.947	0.344	[-0.007, 0.019]
Education	-0.015	-0.284	0.776	[-0.121, 0.090]
Political Orientation	-0.080	-1.543	0.124	[-0.181, 0.022]

Stereotype dispersion: warmth and competence				
Intercept (=Adaptive)	1.888	20.060	.000	[1.703, 2.073]
Exploration Bonus	-0.340	-2.619	.009	[-0.596, -0.085]
Stereotype dispersion: warmth and competence				

Intercept (=Adaptive, Female, Black)	1.941	4.648	.000	[1.130, 2.762]
Exploration Bonus	-0.325	-2.392	.017	[-0.592, -0.058]
Gender (=Male)	0.041	0.294	0.769	[-0.232, 0.314]
Race (=Asian)	0.691	2.395	0.017	[0.124, 1.258]
Race (=Caucasian)	0.423	1.775	0.077	[-0.046, 0.892]
Race (=Hispanic)	0.112	0.340	0.734	[-0.535, 0.759]
Race (=Multiracial)	0.276	0.739	0.460	[-0.458, 1.010]
Age	0.000	-0.017	0.987	[-0.011, 0.011]
Education	-0.041	-0.872	0.384	[-0.134, 0.051]
Political Orientation	-0.072	-1.592	0.112	[-0.160, 0.017]

Note. Estimates are based on an OLS model without (first panels) and with (second panels) covariate variables of age, gender, race, education, and political orientation. *N* = 194 in baseline and *N* = 214 in exploration bonus conditions.

Table S3. Average Treatment Effects (Lower Reward)

	β	*t*	*p* > | *t* |	**[.025, .975]**
Choice entropy: 40-trial current hires				
Intercept (=Adaptive)	2.133	128.688	.000	[2.100, 2.165]
Lower Reward	0.414	17.797	.000	[0.368, 0.460]
Choice entropy: 40-trial current hires				
Intercept (=Adaptive, Female, Black)	2.236	31.890	.000	[2.098, 2.374]
Lower Reward	0.410	17.254	.000	[0.363, 0.456]
Gender (=Male)	-0.026	-1.061	0.289	[-0.073, 0.022]
Race (=Asian)	-0.153	-2.879	0.004	[-0.257, -0.048]
Race (=Caucasian)	-0.178	-4.345	0.000	[-0.259, -0.098]
Race (=Hispanic)	-0.130	-2.174	0.030	[-0.305, -0.037]
Race (=Multiracial)	-0.171	-2.502	0.013	[-0.305, -0.037]
Age	0.001	0.420	0.675	[-0.002, 0.002]
Education	0.011	1.480	0.140	[-0.004, 0.027]
Political Orientation	0.003	0.403	0.687	[-0.012, 0.018]

Choice entropy: 400-trial future hires				
Intercept (=Adaptive)	2.166	75.212	.000	[2.110, 2.223]
Lower Reward	0.356	8.788	.000	[0.276, 0.435]
Choice entropy: 400-trial future hires				

Intercept (=Adaptive, Female, Black)	2.054	16.648	.000	[1.811, 2.296]
Lower Reward	0.353	8.449	.000	[0.271, 0.435]
Gender (=Male)	-0.069	-1.639	0.102	[-0.153, 0.014]
Race (=Asian)	-0.065	-0.695	0.488	[-0.248, 0.119]
Race (=Caucasian)	-0.087	-1.200	0.231	[-0.229, 0.055]
Race (=Hispanic)	-0.063	-0.603	0.547	[-0.269, 0.143]
Race (=Multiracial)	-0.144	-1.199	0.231	[-0.380, 0.092]
Age	0.002	1.179	0.239	[-0.001, 0.006]
Education	0.026	1.866	0.063	[-0.001, 0.052]
Political Orientation	0.014	1.055	0.292	[-0.012, 0.041]

Stereotype dispersion: cooperation and status				
Intercept (=Adaptive)	2.618	23.833	.000	[2.402, 2.834]
Lower Reward	-1.073	-6.953	.000	[-1.377, -0.770]
Stereotype dispersion: cooperation and status				

Intercept (=Adaptive, Female, Black)	2.194	4.626	.000	[1.261, 3.126]
Lower Reward	-1.096	-6.825	.000	[-1.411, -0.780]
Gender (=Male)	0.098	0.599	0.550	[-0.233, 0.418]
Race (=Asian)	0.782	2.181	0.030	[0.077, 1.486]
Race (=Caucasian)	0.348	1.254	0.211	[-0.198, 0.894]
Race (=Hispanic)	0.522	1.295	0.196	[-0.270, 1.314]
Race (=Multiracial)	1.001	2.167	0.031	[0.092, 1.909]
Age	0.007	1.083	0.280	[-0.006, 0.021]
Education	0.041	0.776	0.438	[-0.063, 0.144]
Political Orientation	-0.111	-2.137	0.033	[-0.213, -0.009]

Stereotype dispersion: warmth and competence				
Intercept (=Adaptive)	1.888	20.209	.000	[1.704, 2.071]
Lower Reward	-0.630	-4.796	.000	[-0.888, -0.371]
Stereotype dispersion: warmth and competence				

Intercept (=Adaptive, Female, Black)	1.754	4.340	0.000	[0.959, 2.548]
Lower Reward	-0.657	-4.801	0.000	[-0.925, -0.388]
Gender (=Male)	-0.111	-0.801	0.424	[-0.384, 0.162]
Race (=Asian)	0.734	2.405	0.017	[0.134, 1.334]
Race (=Caucasian)	0.338	1.430	0.154	[-0.127, 0.804]
Race (=Hispanic)	0.588	1.712	0.088	[-0.087, 1.263]
Race (=Multiracial)	0.580	1.475	0.141	[-0.194, 1.354]
Age	0.001	0.182	0.855	[-0.010, 0.012]
Education	-0.003	-0.058	0.954	[-0.091, 0.086]
Political Orientation	-0.049	-1.119	0.264	[-0.136, 0.037]

Note. Estimates are based on an OLS model without (first panels) and with (second panels) covariate variables of age, gender, race, education, and political orientation. *N* = 194 in baseline and *N* = 199 in lower reward conditions.

Table S4. Average Treatment Effects (Random Holdout)

	β	*t*	*p* > | *t* |	**[.025, .975]**
		Choice entropy: 40-trial current hires		
Intercept (=Adaptive)	2.133	127.071	.000	[2.100, 2.166]
Random Holdout	0.333	14.132	.000	[0.286, 0.379]
		Choice entropy: 40-trial current hires		
Intercept (=Adaptive, Female, Black)	2.292	34.362	.000	[2.161, 2.423]
Random Holdout	0.321	13.200	.000	[0.273, 0.369]
Gender (=Male)	-0.049	-2.047	0.041	[-0.096, -0.002]
Race (=Asian)	-0.149	-2.588	0.010	[-0.261, -0.036]
Race (=Caucasian)	-0.176	-4.492	0.000	[-0.253, -0.099]
Race (=Hispanic)	-0.115	-1.857	0.064	[-0.237, 0.007]
Race (=Multiracial)	-0.199	-2.614	0.009	[-0.349, -0.049]
Age	0.001	0.136	0.892	[-0.008, 0.023]
Education	0.001	0.961	0.337	[-0.008, 0.023]
Political Orientation	-0.002	-0.276	0.783	[-0.018, 0.013]

		Choice entropy: 400-trial future hires		
Intercept (=Adaptive)	2.166	72.792	.000	[2.108, 2.225]
Random Holdout	0.180	4.306	.000	[0.098, 0.262]
		Choice entropy: 400-trial future hires		

Intercept (=Adaptive, Female, Black)	2.242	18.668	.000	[2.006, 2.479]
Random Holdout	0.153	3.509	.001	[0.067, 0.239]
Gender (=Male)	-0.029	-0.671	0.503	[-0.114, 0.056]
Race (=Asian)	-0.165	-1.594	0.112	[-0.368, 0.039]
Race (=Caucasian)	-0.181	-2.559	0.011	[-0.593, -0.154]
Race (=Hispanic)	-0.373	-3.341	0.001	[-0.593, -0.154]
Race (=Multiracial)	-0.050	-0.363	0.717	[-0.320, 0.220]
Age	0.001	0.160	0.873	[-0.003, 0.004]
Education	0.029	2.035	0.043	[0.001, 0.058]
Political Orientation	0.001	0.058	0.954	[-0.027, 0.029]

Stereotype dispersion: cooperation and status				
Intercept (=Adaptive)	2.618	24.555	.000	[2.408, 2.827]
Random Holdout	-0.615	-4.107	.000	[-0.909, -0.320]

Stereotype dispersion: cooperation and status

Intercept (=Adaptive, Female, Black)	2.967	6.898	.000	[2.121, 3.813]
Random Holdout	-0.607	-3.877	.000	[-0.915, -0.299]
Gender (=Male)	-0.126	-0.814	0.416	[-0.431, 0.179]
Race (=Asian)	0.879	2.373	0.018	[0.151, 1.606]
Race (=Caucasian)	0.114	0.452	0.652	[-0.383, 0.612]
Race (=Hispanic)	0.314	0.785	0.433	[-0.473, 1.101]
Race (=Multiracial)	0.533	1.084	0.279	[-0.433, 1.499]
Age	0.003	0.047	0.962	[-0.012, 0.012]
Education	0.012	0.238	0.812	[-0.089, 0.114]
Political Orientation	-0.135	-2.638	0.009	[-0.235, -0.034]

Stereotype dispersion: warmth and competence				
Intercept (=Adaptive)	1.888	20.723	0.000	[1.709, 2.067]
Random Holdout	-0.328	-2.562	0.011	[-0.579, -0.076]
Stereotype dispersion: warmth and competence				

Intercept (=Adaptive, Female, Black)	2.189	5.893	0.000	[1.459, 2.920]
Random Holdout	-0.315	-2.327	0.021	[-0.581, -0.049]
Gender (=Male)	-0.161	-1.204	0.229	[-0.424, 0.102]
Race (=Asian)	0.944	2.952	0.003	[0.315, 1.572]
Race (=Caucasian)	0.237	1.083	0.280	[-0.193, 0.666]
Race (=Hispanic)	0.348	1.006	0.315	[-0.332, 1.027]
Race (=Multiracial)	0.114	0.268	0.789	[-0.721, 0.948]
Age	-0.004	-0.761	0.447	[-0.014, 0.006]
Education	0.023	0.525	0.600	[-0.064, 0.111]
Political Orientation	-0.098	-2.230	0.026	[-0.185, -0.012]

Note. Estimates are based on an OLS model without (first panels) and with (second panels) covariate variables of age, gender, race, education, and political orientation. *N* = 194 in baseline and *N* = 200 in random holdout conditions.

Computational Simulations

Formalism: Contextual Multi-Armed Bandit and Bayesian Inference.

Using the job recommendation example in the main text, we consider how a rational agent in a contextual multi-armed bandit setting should solve this problem. We show that strong differences in the allocation of groups to societal positions can be produced by rational agents in the absence of any real inter-group differences and that these agents form stereotype contents along multiple dimensions. The goal of this section is to provide details on the computational modeling approach in the main text. We made a movie (no audio) presentation to animate this mathematical model, readers can watch it in Movie S6.

The agent has access to a discrete number of groups G, and interacts with candidate jobs across discrete trials $t = 1, 2, ..., T$, where the reward is whether or not the job is a good fit for the recommended group member. The jobs are characterized by contextual information $x; x \in \Re^{D}$, that is, jobs have as many as D dimensions of features. Here, in our example, we set D to be 2, corresponding to levels of status and cooperation. For simplicity, we set the number of groups G to be 4, but it can apply to larger finite numbers, as many as the social groups we have in society.

At trial t, the agent observes the current job characterized by its features x_{t}, and the available groups. The goal of the agent is to provide the job with a person from one group who may fit the position. The groups are thus the arms of the bandit, the selection of a group is the action, and the context is the job features x_{t}. After making the recommendation, the agent receives a reward, r_{t}. If the person selected is a good fit for the job, then r_{t} equals 1, if not, r_{t} equals 0. The rewards follow a distribution which can be characterized by the context, x, and parameters, θ_{g}, written $P(r; x, \theta_{g})$ where g is the group to which those parameters correspond. θ is also in $\Re^{D}$, so 2 dimensions in our example. The expected reward for each group, g, can be written as:

$$E\,[r_g \,|\, x,\ \theta_g] \ = \ f(x^T\theta_g), \text{ where } f(.) \ = \ exp(.)\,/\,(1 \ + \ exp(.)).\ [1]$$

where $x^T\theta_g$ is the inner product of these two vectors, being a linear function of x with parameters θ_g. The parameter vector θ_g thus encodes how the dimensions of x, corresponding to the features of the jobs, are weighted for group g when predicting whether a member of that group will be successful in the job. Here, $f(.)$ is a sigmoid function that transforms an arbitrary value into a continuous value in $[0,\ 1]$, to give us $P(r;\ x,\ \theta_g)$.

For $t \ = \ 1,\ 2,\ ...,\ T$, the agent observes past t observations of the contexts, the actions chosen, and their corresponding rewards $(x_t,\ g_t,\ r_t)$. Importantly, no payoff information is revealed for the unchosen groups, $g \ \neq g_t$. The objective is to find a solution that minimizes the cumulative regret; the regret is the expected difference between the optimal reward received by always playing the optimal group, g_t^*, and the reward received by following the actually chosen group, g_t. Thus, the cumulative regret at the end of the game, $R(T)$ can be written as:

$$R(T) \ = \ \sum_{t=1}^{T} E[r_{g^*}|x_t, \theta_{g^*}] \ - \ E[r_g|x_t, \theta_g].\ [2]$$

Finding the optimal solution to this problem requires balancing between exploration and exploitation. While there are no known optimal solutions to contextual bandits, we focus on an approach known as Thompson sampling (*26, 27*) which generalizes an optimal solution to the standard multiarmed bandit. Thompson sampling has previously been used to show how adaptive exploration can produce stereotypes in a simpler context-free multiarmed bandit setting (*29*) and has been shown to be a good model of human choices on contextual bandit tasks (*30*).

Using the same job recommendation example, Thompson sampling for contextual bandit can be defined in terms of the Bayesian solution to the problem of estimating θ_g. For each group, g, if we know θ_g, then applying any context x_t, we can derive the expected reward via Equation 1. But we do not know the parameters θ_g, so the goal is to estimate them. At time step t, first, a prior distribution $P(\theta_g)$ represents uncertainty over the parameter space and the likelihood function $P(r_g|x_t,\ \theta_g)$ represents the probability of reward given a context $x_t$and a parameter θ_g. Applying Bayes' rule, the posterior distribution over θ_g is given by:

$$P(\theta_g|r_g, x_t) \propto P(r_g|x_t, \theta_g)\ P(\theta_g).\ [3]$$

The posterior distribution, therefore, represents the updated beliefs about the parameters $\theta_g$after incorporating the new evidence and the prior belief. Next, a sample $\theta_{t+1,g}$is randomly drawn from this posterior, corresponding to a stochastic estimate of θ_g after t time steps. The agent follows this procedure – estimate θ_g, draw a random sample $\theta_{t+1,g}$ – for all groups, and plays the group for which the predicted probability of reward is highest. This is equivalent to sampling each group with a probability corresponding to the posterior probability that group is most likely to generate a reward, which is Thompson sampling.

Specifically, in Equation 3 we assume the prior, $P(\theta_g)$, follows a Gaussian distribution $N(\mu_0, S_0)$ and the likelihood, $P(r_g|x_t, \theta_g)$, follows a Bernoulli distribution, with the joint probability mass function over the rewards:

$$\prod_{t=1}^{T} P(r_t = 1|x_t, \theta_t) = \prod_{t=1}^{T} [1\ /\ (1\ +\ e^{-\theta_t^T x_t})]^{r_t}\ [e^{-\theta_t^T x_t}\ /\ (1\ +\ e^{-\theta_t^T x_t})]^{1-r_t}\ .\ [4]$$

The posterior distribution derived from this joint probability distribution is intractable, hence, we use Laplace's method to approximate the posterior distribution with a multivariate Gaussian distribution with a diagonal covariance matrix. The mean of this distribution is the maximum-a-posteriori estimate, and the inverse variance of each feature is the curvature (Algorithms 3 in *28*).

Simulation Results for Main Hypothesis:

In this section, we present predictions derived from the above model with simulation data. Again, for all simulations, we use Bayesian logistic regression to estimate the function between job features and groups and use Thompson sampling to make decisions about how to solve the explore-exploit dilemma. As defined above, the context vector has two dimensions, corresponding to status and trust with binary features: $\{1, 1\}$ indicates high status and high trust jobs such as doctors, $\{1, -1\}$ indicates high status low trust jobs such as lawyers, $\{-1, 1\}$ indicates low status high trust jobs such as childcare aides, and $\{-1, -1\}$ indicates low status and low trust jobs such as janitors. There are four groups; whose reward distributions are independent from each other. The current model has the same intercept for all groups as we assume no group-level differences. The underlying expected reward probability centers around 0.9, $N(0.9,\ 0.001)$, for all groups. We made the variance small because we assume all groups are equally and highly likely to be successful. The agent starts with a prior belief follows a normal distribution of $N(0,\ 1)$. With this set up, we ran 100 simulations. Within each simulation, the Bayesian agent played 40 rounds of the game.

The critical prediction from our model is that the Bayesian agents will end up creating a biased social structure such that certain groups are selectively recommended to certain jobs, compared to that produced by agents who make choices at random. Two key outcome variables quantify this hypothesis: selective recommendation patterns and dispersed mental representations.

First, for recommendation choices, we predict Bayesian agents do not recommend jobs indifferently, but rather should differentially recommend certain groups to do certain kinds of jobs. This occurs because of the explore-exploit tradeoff: Having found a group that performs well at a given job, searching for other groups that might also perform well is costly, and it is better to focus on the group that is known to perform well. However, this selective choice should not appear in random decisions when the agents do not intend to use past success experiences to solve the explore-exploit tradeoff. One way to quantify the randomness of a system is entropy

(Shannon, 1948). Given an output 4-by-4 matrix N where the rows represent groups and the columns represent jobs, with a number of assignments $n_{g,j}$ in each cell:

$$H(N) =- \sum_{g,j} n_{g,j}/n \log n_{g,j}/n \text{ . [5]}$$

where n is the total number of assignments. We can compare this entropy value between choices made by the Bayesian agents and the random-decision agents.

Second for mental representations, we predict Bayesian agents will develop dispersed mental maps for the four groups along the two dimensions as a result of these differential recommendations. Here, we use the estimated coefficient vector θ to approximate the agents' mental model of each group's perceived trustworthiness and competence. The Bayesian agents should give differential estimates of the parameters given their selective experiences, whereas the random-decision agents should give relatively equal estimates of the parameters given they encounter similar amounts of experiences with all groups. Given a learned two-dimensional array of coefficients, we can calculate the summed Euclidean distance S among the four groups:

$$S(\Theta) = \sum_{g} \sqrt{\sum_{d} (\theta_{g,d} - \mu_{d})^{2}} \text{. [6]}$$

where Θ refers to the collection of all θ_g, θ_g refers to the estimated coefficients for each group, and μ is the averaged coefficients for all groups. We can compare the mental representation distance of the estimated coefficients between the Bayesian agents and the random-decision agents.

Below we present results of Equations 5 and 6 from the Bayesian agents and the random-decision agents. To emphasize, the ground truth represents an original egalitarian social world: among 10 potential pairs of jobs and groups, approximately 9 pairs generate a positive reward of 1 and only 1 pair generates a reward of 0. As a natural consequence, the most accurate mental map corresponding to this original social world should position groups close to each other in terms of contextual features.

First, the random-decision condition provides a sensible baseline; take one simulation as an example (Fig. S6). When randomly exploring the world, the agent recommended approximately equal numbers of each job to each group (Fig. S6a). Because of relatively equal allocations of jobs and groups, we did not observe consideration distances among the learned weights among the four groups in this random-decision agent (Fig. S6b). This implies that the simulated random-decision agents did not form specific stereotypes of the four arms/groups.

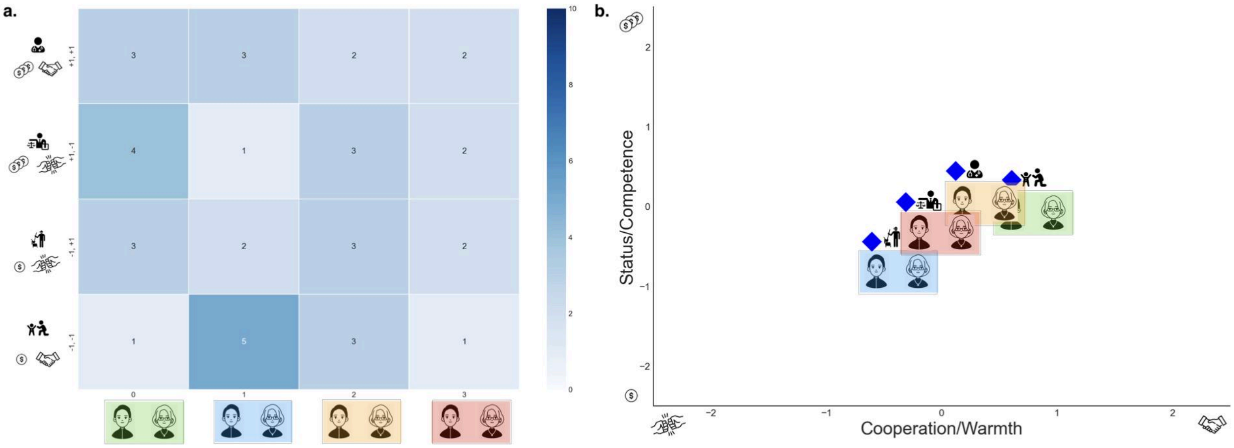


Fig. S6. An example simulated results from an agent who makes decisions at random. The heatmap on the left shows how many times a group (on the horizontal axis) is recommended for a job (on the vertical axis). The scatterplot on the right shows estimated coefficients for the four groups on two binary features.

Next, the Bayesian decision condition provides our critical prediction; take one simulation as an example (Fig. S7). This simulated Bayesian agent confirmed the intuition given in the introduction. Instead of recommending groups equally to jobs, the agent selectively recommended one particular job to mostly one group, 9 or 10 times, and was, therefore, less likely to recommend the other three jobs to the same group, 1 or 2 times (Fig. S7a). As a result of such selective recommendation, we saw considerable variation in the estimated weights, such as associating one group strongly with one feature or another group with another feature (Fig. S7b). The dispersed mental representation indicates the emergence of stereotypes.

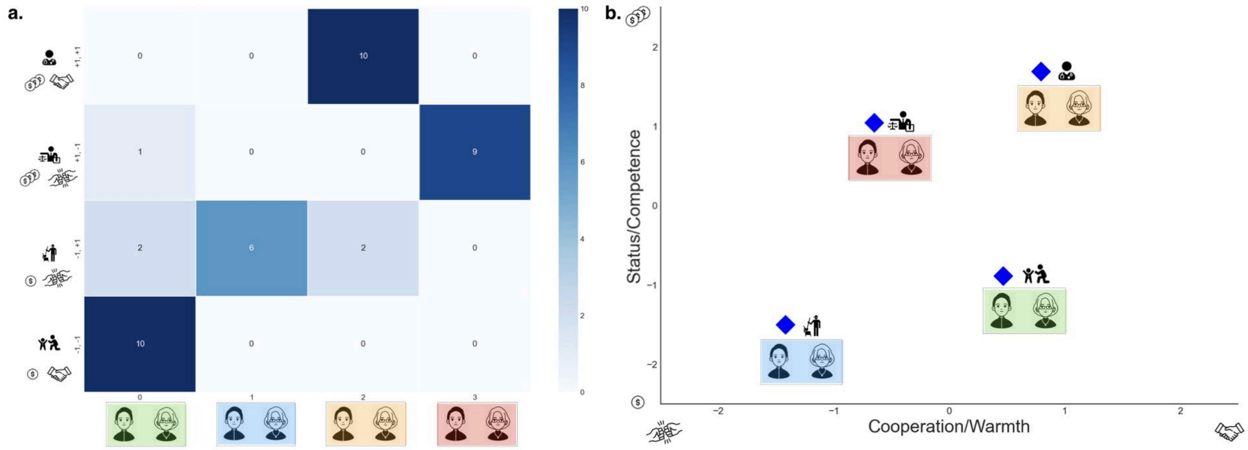


Fig. S7. An example simulated results from an agent who makes adaptive decisions using Thomson sampling. The heatmap on the left shows how many times a group (on the horizontal axis) is recommended for a job (on the vertical axis). The scatterplot on the right shows estimated coefficients for the four groups on two binary features.

Moving beyond individual examples, we next compared the aggregate-level pattern across 100 simulations. To compare across simulations while also preserving each simulation's characteristics, we rank-ordered the choices within each simulation. The results confirmed the individual examples: On average, Bayesian agents were more likely to selectively recommend jobs to different groups (Fig. S8a) as compared to random-decision agents (Fig. S8b).

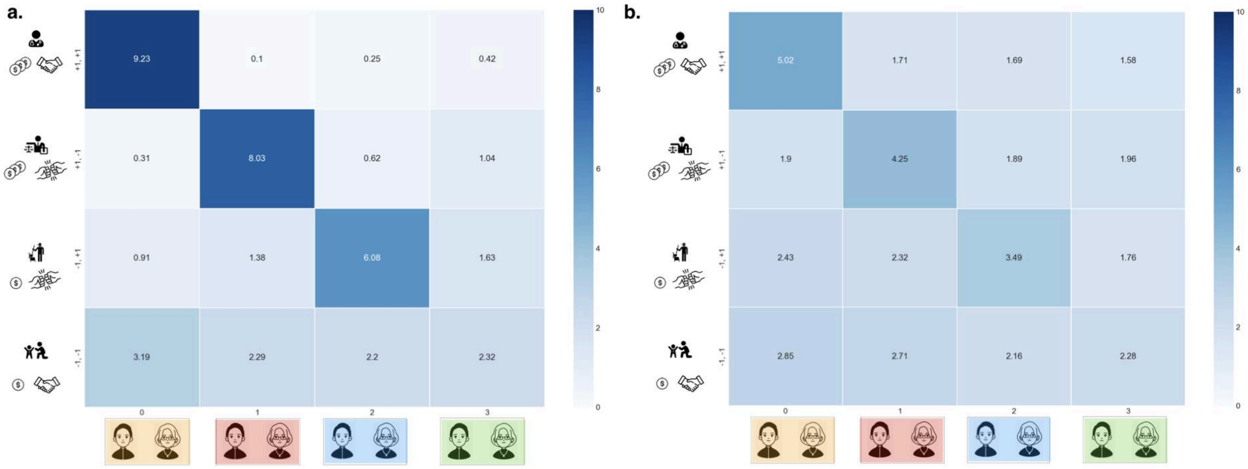


Fig. S8. Heatmaps between the two conditions, aggregated across 100 simulations after rank order. The reason for rank order is that each simulation starts with a different prior, by chance, and therefore, different subsequent decisions. Simply averaging across simulations loses the

original features within each simulation. To minimize the loss of information, we rank-ordered each simulation. Specifically, for each simulation, we find the max value of the entire 4-by-4 matrix, and store that row and column as the first row and column. We then find the max value of the remaining 3-by-3 submatrix, store that row and column as the second row and column, and repeat the same procedure for the remaining submatrices. After this transformation, we obtained an aggregate summary for which the first row and the first column always store the max value, the second row and column always store the second max value, etc.

To examine the robustness of this descriptive result, we ran statistical analyses across 100 simulations between the random decision condition and the Bayesian decision condition. We used an Ordinary-Least-Square linear regression model with the condition as the predictor variable (Bayesian coded as 0 vs. random coded as 1), choice entropy (Eq. 5), and mental map dispersion (Eq. 6) as the outcome variable. We found the Bayesian condition showed a smaller entropy (treatment effect: $b = 0.645$, 95% *CI* [0.614, 0.676], $p < .001$) and a bigger dispersion (treatment effect: $b = -1.447$, 95% *CI* [-1.596, -1.297], $p < .001$) than the random-decision condition. In other words, this result confirmed the above descriptive analysis: agents who use their past success to guide new decisions to solve the explore-exploit dilemma were more likely to differentially allocate groups and to form dispersed mental maps than agents who make decisions at random.

Simulation Results for Mechanism/Interventions:

Here we provide details on the intervention simulations. Specifically, we simulated three interventions that are hypothesized to diversify choices and reduce stereotypes.

First is the exploration bonus. This is a mechanism that is commonly used to support exploration by reinforcement learning systems in computer science. According to one popular method for creating an exploration bonus, known as count-based exploration, we count how many times a state (group-job pair) has been encountered and assign a bonus accordingly (Bellemare et al., 2016). The bonus guides the agent's behavior to prefer rarely visited states to common states. Let $N_n(s)$ be the empirical count function that tracks the real number of visits of a state $s$in the sequence of $s_{1:n}$. The bonus is then proportional to $\sqrt{1/1 + N_n(s)}$. For example, if Tufa has been selected twice, the bonus reward will be $\sqrt{1/1 + 2}$ = 0.577, and if this time, Tufa is a good choice, the base reward is 1, therefore the total reward will be 1.577 for choosing Tufa. However, if Aima has not been selected at all, the bonus reward will be $\sqrt{1/1}$ = 1, and if this time, Aima is a good choice, the base reward is 1, therefore the total reward will be 2 for choosing Aima. The optimal solution is to choose Aima instead of Tufa, which can increase exploration. See Code S4 Exploration Bonus and Figs. S9a and b.

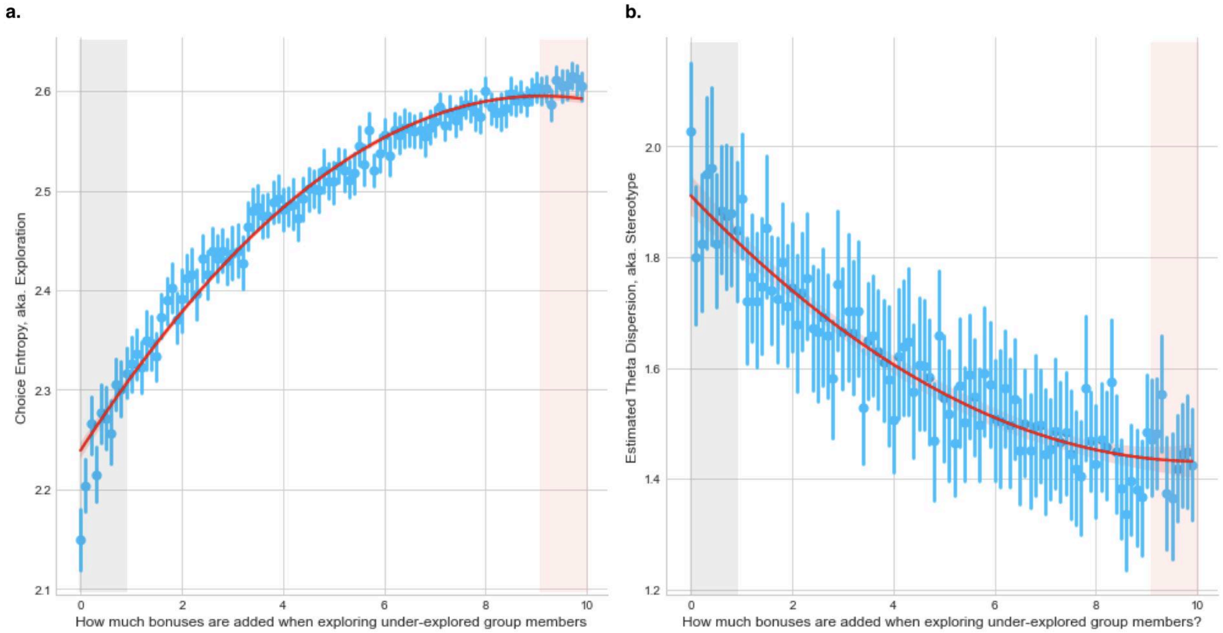


Fig. S9. Exploration bonus intervention. Panel a shows the increase in choice entropy as a function of the unit price of the expiration bonus, that is, the more you pay for exploration, the more diversified choices you will see. Panel b shows the decrease in stereotype dispersion as a function of the unit price of the exploration bonus, as a consequence of increased exploration. Grey bars highlight baseline conditions whereas red bars highlight the intervention conditions that we use to design human experiments.

The second intervention is to make the tasks more challenging. In the baseline model, we used an expected reward of 0.9 as the ground truth, making it very likely for the players to get a reward. However, we can also decrease the expected reward. As a consequence, players are more likely to encounter negative experiences which in turn can encourage them to explore new options. See Code S4 Challenging Tasks and Figs. S10a and b.

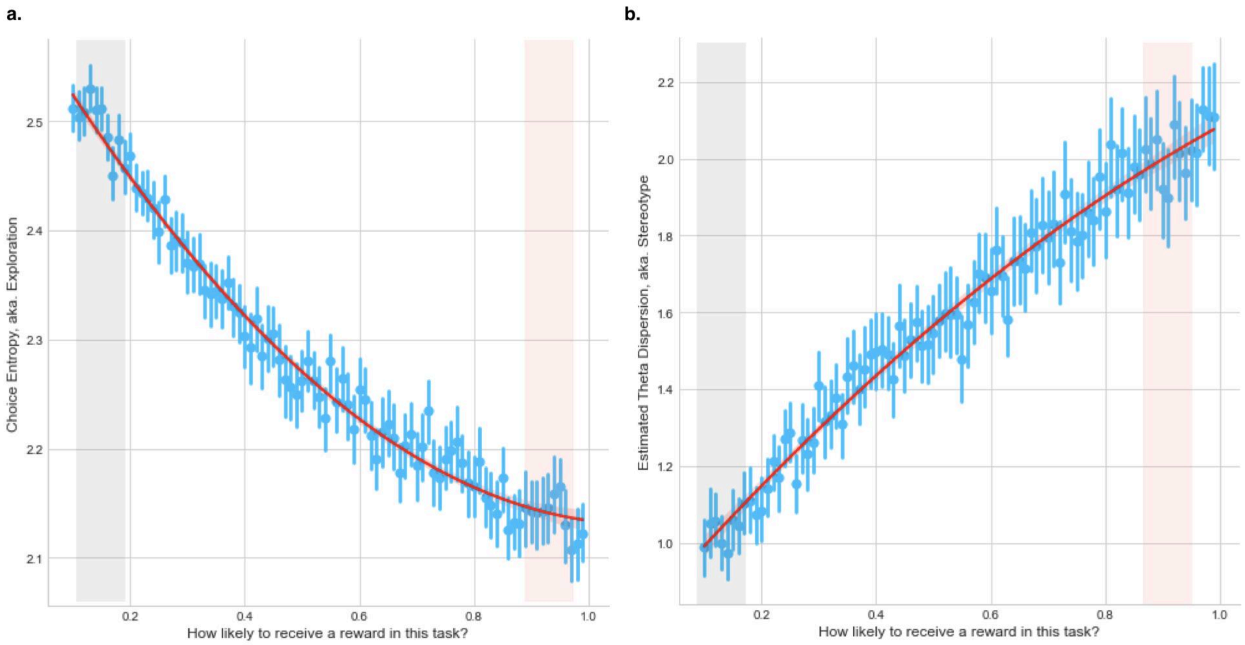


Fig. S10. Expected reward intervention. Panel a shows the decrease in choice entropy as a function of the expectation of getting a reward in the game, that is, the less likely you think you will get a reward, the more diversified choices you will make. Panel b shows the increase in stereotype dispersion as the expected reward increases. Grey bars highlight baseline conditions whereas red bars highlight the intervention conditions that we use to design human experiments.

The third intervention is to make some groups unavailable, at random. When agents make decisions, they can always choose from all groups, so if they want, they can always stick to their known options. However, if some groups are unavailable, the structure forces the agents to explore other options. We varied the rate of unavailability and simulated the intervention effects. That is, in some conditions, 10% of the trials will make two out of four groups unavailable, but in other conditions, 50% of the trials will make two out of four groups unavailable, or other times, the rate is 90%. See Code S4 Holdout At Random and Figs. S11a and b.

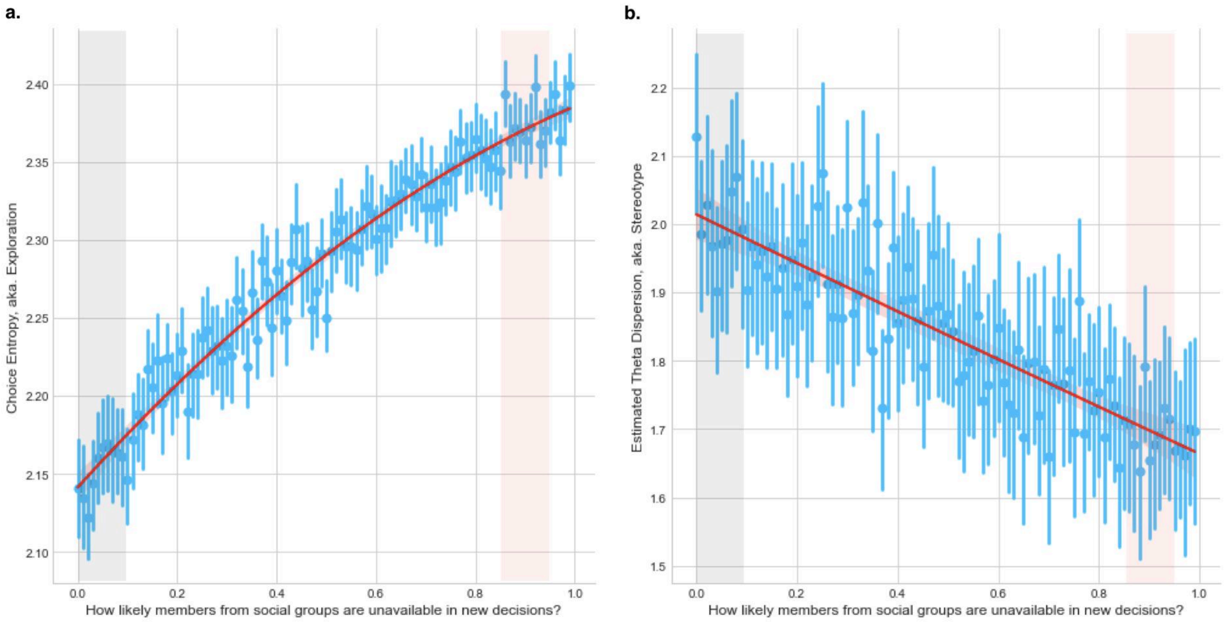


Fig. S11. Random holdout intervention. Panel a shows the increase in choice entropy as a function of the likelihood two out of four groups are unavailable when agents need to make a decision. That is, the more likely you see two groups, at random, are unavailable, the more likely you explore other groups. As a result, Panel b shows a decrease in stereotype dispersion as the unavailability increases. Grey bars highlight baseline conditions whereas red bars highlight the intervention conditions that we use to design human experiments.

Simulation Results for Other Variants:

In the main text, we designed parameters to reflect our theoretical claims. In particular, we decided to fix the underlying ground truth to be high and identical for all combinations of contextual features and all groups. This is to minimize the confound of stereotype accuracy. We found even if all groups are equally rewarding, the adaptive decision agents were unable to recover that truth. Nonetheless, some readers may be interested in what might happen when the ground truth indeed differs. Here we present the simulation results for this variant.

In simulations where the true reward distribution is identical, as follows ($\theta = .9$):

	Group 1	Group 2	Group 3	Group 4
[1,1]	95	87	93	89
[1,-1]	88	92	92	82
[-1,-1]	91	92	92	88
[-1,1]	91	92	89	91

The Thompson sampling agents decide as follows:

	Group 1	Group 2	Group 3	Group 4
[1,1]	1	94	5	0
[1,-1]	87	1	0	12
[-1,-1]	0	0	100	0
[-1,1]	2	0	1	97

In simulations where the true reward distribution is different, as follows ($\theta = .9$ vs .1):

	Group 1	Group 2	Group 3	Group 4
[1,1]	88	10	12	9
[1,-1]	16	87	14	13
[-1,-1]	8	14	92	10

[-1,1]	6	12	7	87

The Thompson sampling agents decide as follows:

	Group 1	Group 2	Group 3	Group 4
[1,1]	93	6	0	1
[1,-1]	1	95	1	3
[-1,-1]	1	0	99	0
[-1,1]	2	0	3	95

In simulations where the true reward distribution is different, slightly, as follows (θ =. 9 vs. .8):

	Group 1	Group 2	Group 3	Group 4
[1,1]	92	86	79	84
[1,-1]	82	91	82	74
[-1,-1]	81	81	89	77
[-1,1]	90	83	81	86

The Thompson sampling agents decide as follows:

	Group 1	Group 2	Group 3	Group 4
[1,1]	0	0	0	100
[1,-1]	0	0	99	1
[-1,-1]	1	99	0	0
[-1,1]	98	1	1	0

In sum, we found that when the ground truth indeed differs significantly (0.9 vs. 0.1), the adaptive decision agents can recover that difference. However, when the differences are not that big (0.9 vs. 0.8), the adaptive-decision agents behave as if they recovered some differences, which significantly exaggerated the ground truth difference.