Learning and Information Use in an Intergroup Context

Benjamin G. Mittman (bgm37@case.edu)
Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine

Keelah E.G. Williams (kewillia@hamilton.edu)
Department of Psychology, Hamilton College

Ting Qian (ting.qian@findingfive.com)
FindingFive Corporation

Vikranth R. Bejjanki (bejjanki@hamilton.edu)
Department of Psychology and Program in Neuroscience, Hamilton College

Abstract

When faced with uncertainty, human observers maximize performance by integrating sensory information with learned task-relevant regularities. Does this behavior similarly occur in social settings? In this paper, we explore how reward-seeking behavior in an intergroup context is affected by readily available but task-irrelevant social information (in the form of group membership) when task-relevant reward information can be learned over time. Across two experiments, we show that participants learned and utilized task-relevant regularities to inform their choices. We also show that human observers are not universally biased towards utilizing social information in all settings—participants learned to disregard social information when not relevant to the task at hand. However, learning about the utility of social information (Experiment 2) had a long-term influence on observers' ability to subsequently learn and utilize available sources of information. Real-world intergroup contexts typically encompass situations and stimuli that have been previously experienced by the observer. Our findings highlight the powerful influence of learning in such contexts.

Keywords: statistical learning; intergroup interactions; judgment and decision making

Introduction

The human brain must manage ambiguous or uncertain information at every level of analysis—e.g., due to noise in sensory measurements or variability in the environment (Knill & Pouget, 2004). Yet human observers are adept at mitigating the influence of such uncertainty on their behavior by effectively utilizing available sources of information (Bernardo & Smith, 2008; Cox, 1946). For instance, they can improve task performance by utilizing relevant information extracted from the environment using statistical learning techniques (Reber, 1989). In a wide range of tasks, including visuo-motor integration (Körding & Wolpert, 2004), timing behavior (Jazayeri & Shadlen, 2010), cue combination (Adams et al., 2004), categorical judgments (Huttenlocher et al., 2000), and movement planning (Hudson et al., 2007), observers learned task-relevant environmental statistics and combined this knowledge with the uncertain information available on each trial, thus improving performance. Notably however, much of this prior work has examined human behavior in nonsocial domains. To what extent do such mechanisms also influence behavior in social settings?

Human observers act as biased information processors in social settings, utilizing some types of information more than others (e.g., Averbeck & Duchaine, 2009; Sharot et al., 2011; for reviews see Amodio & Mendoza, 2010; Hewstone et al., 2002). For instance, they rely on group-based expectancies when evaluating individuals, even when this information may not be task-relevant (Devine, 1989; Hamilton & Sherman, 1994; Spencer et al., 1998). Such biases may be adaptive in some situations. For instance, the brain is adapted to identify and manage the threats and opportunities afforded by others in one's social environment (McArthur & Baron, 1983; Neuberg et al., 2011; Zebrowitz & Montepare, 2006). Such identifications must often be based on salient but uncertain social cues (Blascovich et al., 1997). One salient cue to potential threat is outgroup membership. Intergroup conflict and aggression have been recurrent features of human social life (Chagnon, 1988; Ferguson, 1984). To the extent that cues of group membership and affordance-relevant behaviors (e.g., confrontation) reliably covaried, associating outgroup members with traits connoting danger would have promoted adaptive behavior. One would therefore expect observers to possess psychological mechanisms that are sensitive to cues of group membership (Kurzban et al., 2001) and that motivate behavior consistent with ingroup favoritism (Brewer, 1979, 1999). Yet, an over-reliance on social information carries potential costs. Although social information is task-relevant in some settings (Jussim et al., 2015), an automatic categorization of threat in situations where social information is irrelevant (Dovidio et al., 2002) would result in an inappropriate implicit bias favoring the ingroup over the outgroup (Dasgupta, 2004; Everett et al., 2015). Thus, observers should be flexible, capable of engaging in a deeper individuation of targets and a reweighting of available sources of information if heuristics fail to sufficiently capture the likely affordances of the target or when the outcome depends on greater accuracy (Fiske & Neuberg, 1990; Pendry & Macrae, 1994). In such situations, learning is a critical mechanism for appropriately adjusting one's behavioral responses (Neuberg et al., 2011).

Previous work examining the use of social information in decision-making has often considered situations in which participants have to make single-shot intergroup decisions (e.g., allocating resources to different groups; Tajfel, 1970).

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

This approach makes it difficult to determine the role of learning. For instance, would participants continue to utilize social information in situations in which they have the opportunity to learn that it is irrelevant to the task at hand? Indeed, although social information is readily available in most social settings, there is often uncertainty about the extent to which it is task-relevant. In single-shot intergroup decisions, given such uncertainty, participants may be biased towards drawing on their prior knowledge about the utility of social information (e.g., group membership) and rely on this source of information when making decisions. In contrast, given the opportunity to learn the task-relevance of available sources of information, an open question is whether participants continue to be biased towards utilizing social information despite learning that it is not task-relevant.

In the current study, by repeatedly exposing participants to a decision-making task involving social information, we examined whether and how human observers extracted task-relevant information via learning, and how social information (group membership) interacted with learned information (reward contingencies). As a first step, we considered the situation in which a few exposures to the task were sufficient to learn that social information was not task-relevant.

Experiment 1

Participants carried out a probabilistic decision-making task in which they selected, on each trial, which of two faces (one Black and one White; drawn from a set of eight Black and eight White faces) would maximize their rewards. Unbeknownst to them, half of the Black and half of the White faces were associated with a greater probability of reward—group membership was therefore not diagnostic for the task of deciding which of the two faces presented on each trial was more likely to be rewarded. Crucially, participants were also given the opportunity to learn the task-relevant reward information—which of the Black and White faces were more likely to be rewarded—over repeated exposure to the task.

Method

Participants Sixteen self-identified White undergraduates (8 women, M_{age} = 20.4, SD_{age} = 1.67) completed the study. Informed written consent was obtained from all participants, who were compensated $6 per half-hour of participation.

Materials Face images were obtained from the Chicago Face Database (Ma et al., 2015), a multiracial database with high-resolution color images and independent ratings for several facial features. Individual faces were selected to control for key features (identified *a priori*) which could have influenced their evaluation by participants, including age, emotional expressiveness, and traits related to avoidance behavior. In line with the literature on intergroup contexts (see Amodio & Mendoza, 2010), only male faces were used to further reduce variability in the normed dimensions of the faces, especially threateningness (Wilson et al., 2017). Eight White and eight Black faces meeting these inclusion criteria were randomly selected and designated as “ingroup” and “outgroup,” respectively. All images were retouched to remove any hair blocking the eyes, cropped to remove excessive background surrounding the faces, and scaled to 320 x 360 pixels.

Procedure Participants completed a task inspired by a prior study examining the influence of emotional expression cues on decision-making (Averbeck & Duchaine, 2009). The task consisted of 10 blocks of 128 trials each, displayed using MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997). For each participant, half the ingroup and outgroup faces were randomly assigned to a high-probability reward group (labeled “group A”) and were associated with a reward 70% of the time. The remaining faces were assigned to a low-probability reward group (labeled “group B”) and were associated with a reward 30% of the time. This design resulted in an equal probability (on average) of reward for ingroup and outgroup faces (i.e., social information did not predict reward). On any given trial, a group A face could only be paired with a group B face and an ingroup face could only be paired with an outgroup face. Within each block, each face was shown a total of 32 times in pseudorandom order (16 times on each side of the screen).

Participants began with a base payment of $5 and were told that they could earn a bonus based on their performance—five points for each “correct” (i.e., rewarded) decision, with 100 points translating to a bonus of 10 cents (all participants were nonetheless paid at the same rate—greater than the maximum possible amount they could earn—but were not informed of this until after the study). After task completion, each participant was probed to ensure they were initially naïve to the study and properly understood the task.

Each trial consisted of three phases (Fig. 1): A blank gray screen was initially presented for 1s. Two images were then presented on either side of the screen for up to 5s. If a response was not made within 5s, participants were prompted to respond more quickly next time, and the missed trial randomly reappeared within the current block. If a response was made within 5s, feedback lasting 1s was presented. Feedback included the correct image for the current trial (regardless of the participant’s response), either a 2.5-kHz “win” tone for a correct response or a 1.25-kHz “lose” tone for an incorrect response, and text informing participants whether they won or lost.

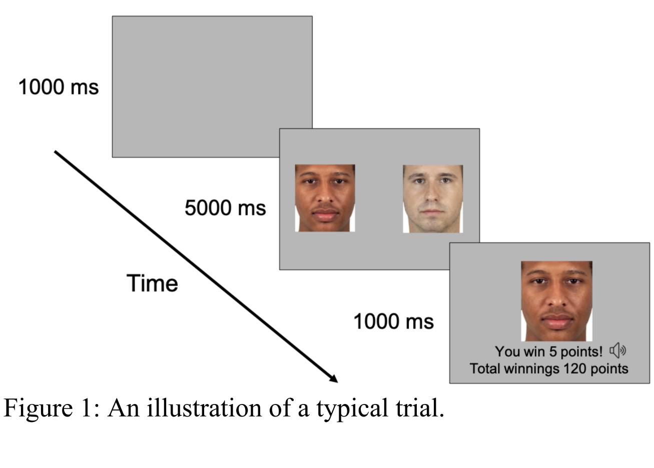

Figure 1: An illustration of a typical trial.

Analyses Data analysis was performed using R (v4.2.2; R Core Team, 2022). Participants' choices were modeled as the outcome variable using mixed effects logistic regression (Jaeger, 2008), implemented via the lme4 package (v1.1.31; Bates et al., 2015). Participants' use of reward versus social information was modeled separately, with group A and ingroup coded as successes, respectively. Trial number (as a continuous variable) and block number (as a categorical variable) were included as fixed effects. To account for randomness in learning rates across participants, a by-participant random slope of block number was included. We fit each regression model under two variations: one with an intercept and one without. In the second "intercept-free" variation, the removal of the intercept allows an intuitive interpretation of the coefficients of block number as participants' block-level performance against chance (i.e., a log-odds of 0). In the more standard variation, the coefficients of block number indicate differences in participants' block-level performance as compared to that in the first block, which reflects *learning* due to increased exposure. All regression coefficients are reported as raw log-odds.

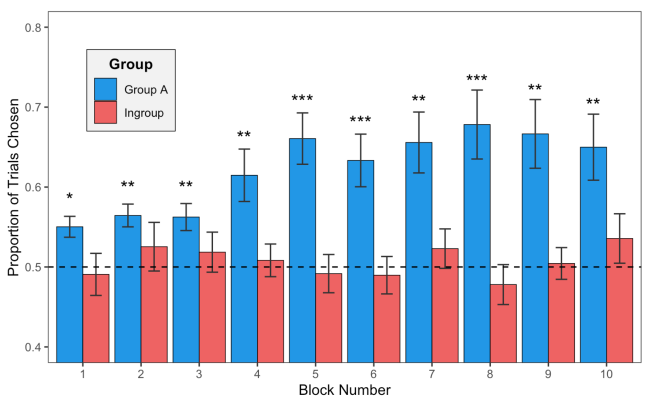

Figure 2: Learning and information use in Experiment 1. Error bars reflect standard errors of the mean; ***$p < .001$, **$p < .01$, *$p < .05$.

Results

Participants exhibited a robust preference for faces belonging to the high-probability reward group (group A) over faces belonging to the low-probability reward group (group B) in every block of Experiment 1 (ß range [0.15, 0.88], all $p \leq .01$; Fig. 2). Notably, this preference developed rapidly—we found a reliable preference for group A within the first block (ß = 0.15, p = .012). Participants' preference for faces in group A increased across blocks, reflecting greater use of the reward information as a function of exposure to the task. Compared to block 1, the log-odds of choosing group A was significantly higher in blocks 4-10 (ß range [0.30, 0.73], all $p < .04$). Conversely, we found no evidence for a preference for either ingroup or outgroup faces in any block (ß range [-0.16, 0.12], all $p > .14$; Fig. 2). The use of social information also did not vary as a function of task exposure: the log-odds of choosing ingroup was not significantly different in blocks 2-10 (ß range [-0.06, 0.22], all $p > .18$) than in block 1. These results suggest that participants rapidly learned and utilized the task-relevant reward information, while disregarding the available but irrelevant social information.

Experiment 2

In Experiment 1, participants were presented with a situation in which social information was readily available whereas reward information needed to be learned over time. At the outset of the task, participants were unaware that social information was task-irrelevant. However, repeated exposure allowed them to learn the task-relevant reward contingencies, including that social information was irrelevant to the specific task at hand. We found that participants rapidly learned the task-relevant reward information and showed increasing use of this information over time. Furthermore, participants' behavior did not reflect a distinguishable use of social information at any point in the experiment. These results are therefore consistent with the notion that—unlike in situations involving single-shot decisions, where participants draw on their prior knowledge about the utility of group membership and demonstrate a preference for the ingroup—having the opportunity to learn about the irrelevance of social information allowed participants to flexibly disregard it. However, this pattern would also be consistent with the possibility that our participants simply approached the task without a prior preference for ingroup faces. Indeed, although an implicit preference for the ingroup has been widely documented, recent studies have suggested that it may be less common among young people (Nosek, 2007), and that its overall magnitude may have decreased during the past decade (Charlesworth & Banaji, 2019).

To address this possibility, we ran a follow-up experiment that differed in two key ways. First, the task was divided into two parts. In part 1 (comprising three blocks), social information was made task-relevant such that faces were rewarded based on their group membership. To accomplish this, the reward probabilities used for group A and group B in Experiment 1 were now applied to the ingroup and outgroup categories. Subsequently, in part 2 (comprising six blocks), the reward and social information were distributed identically to Experiment 1. That is, groups A and B were assigned reward probabilities of 70% and 30% respectively, and ingroup and outgroup faces were rewarded equally. The second modification was the creation of two conditions: ingroup preference (IP) and outgroup preference (OP). Ingroup faces were rewarded 70% of the time and outgroup faces 30% of the time for IP, and vice versa for OP.

In part 1, based on the results from Experiment 1, we expected that participants would rapidly learn the task-relevant reward information (i.e., that group membership predicted reward), and that their behavior would therefore reflect a preference for the social category that was associated with the higher probability of reward. Furthermore, if this prediction was supported, we would be guaranteed that at the outset of part 2, participants would exhibit a preference for either the ingroup or outgroup (depending on the condition). As such, replicating the first six blocks of Experiment 1 in part 2 of Experiment 2 would allow us to examine how reward and social information interact given a pre-existing preference for one of the social categories.

Method

Participants 33 self-identified White undergraduates (14 women, M_{age} = 19.9, SD_{age} = 1.29) took part in Experiment 2. Participants were randomly divided into two conditions; 17[1] were assigned to the ingroup preference (IP) condition and 16 to the outgroup preference (OP) condition. Informed written consent was obtained from all participants, who were compensated $6 per half-hour of participation.

Materials Stimuli used in part 1 consisted of a set of eight White and eight Black faces that fit the criteria for (but were not used in) Experiment 1. Stimuli used in part 2 were the same White and Black faces that were used in Experiment 1. Different faces were used in parts 1 and 2 to ensure that patterns of information use in part 2 could be attributed to effects of group membership and not to individual faces.

Procedure The procedure for Experiment 2 was similar to that used for Experiment 1, except that participants were told before the start of the experiment that the computer task consisted of two segments with a break between them.

Analyses As with Experiment 1, mixed effects logistic regression was used to model participants' choices. In part 1, the use of reward information and social information was modeled separately, with group A and ingroup (IP condition) or outgroup (OP condition) coded as successes, respectively. IP and OP conditions were modeled separately. In each model, trial number and block number were included as fixed effects, with by-participant random intercepts and by-participant random slopes of block number. Intercept-free variations of the models were also fit to compare participants' block-level performance against chance. Finally, the IP and OP conditions were combined into one model, with condition modeled as an interaction between trial number and block number, to directly compare their learning rates. The same models described for part 1 of Experiment 2 were replicated for part 2, with the addition of a combined model to test the interaction effect of learning rates between conditions. Additionally, we compared participants' choices for group A/B between experiments with mixed effects models as described above, but with experiment as an interaction term to compare learning between the two experiments.

Results

Part 1 As expected, IP participants exhibited robust preferences for ingroup faces (ß range [0.42, 0.76], all $p < .002$; Fig. 3A) and OP participants exhibited robust preferences for outgroup faces (ß range [0.70, 1.28], all $p < .001$; Fig. 3B) in each of the 3 blocks of part 1. Furthermore, IP participants' preference for ingroup faces increased across blocks; compared with block 1, they were significantly more likely to choose ingroup faces in block 2 (ß = 0.25, $p = .002$) and block 3 (ß = 0.36, $p < .001$). Similarly, OP participants showed marginal increases in their preference for outgroup faces in block 2 (ß = 0.34, $p = .054$) and block 3 (ß = 0.58, $p = .06$), compared to block 1. Notably, we did not observe a reliable difference between the two conditions in the rate at which participants' preferences changed across blocks (ß = 0.28, $p = .11$). Finally, neither IP nor OP participants showed reliable preferences for the arbitrary groups A/B at any point in part 1 (ß range [-0.10, 0.04], all $p > .12$), nor did this change across blocks (ß range [-0.14, 0.02], all $p > .08$), verifying our manipulation check.

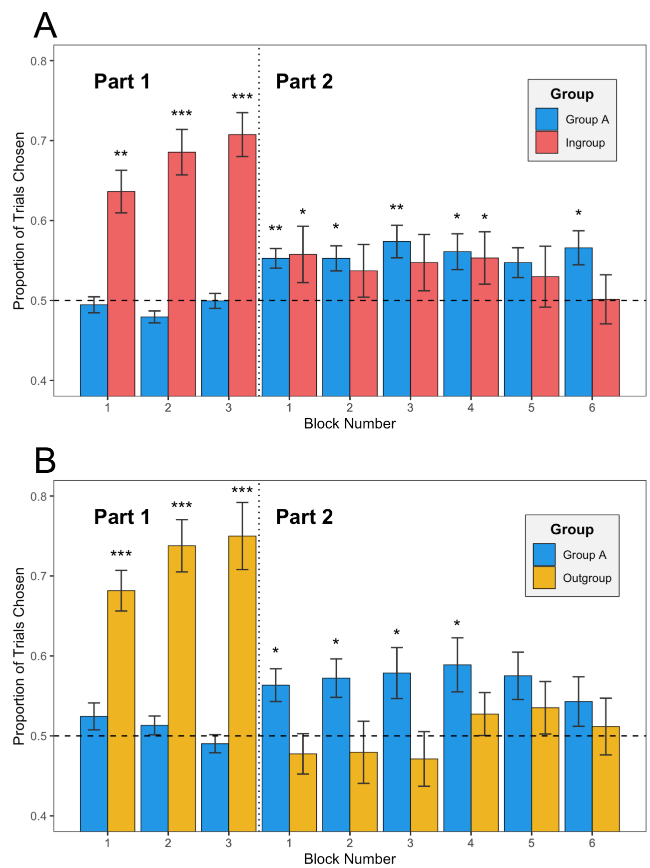

Figure 3: Learning and information use in Experiment 2, IP (panel A) and OP (panel B) conditions. Error bars reflect standard errors of the mean; ***$p < .001$, **$p < .01$, *$p < .05$.

Part 2 IP participants showed a robust preference for faces belonging to the high-probability reward group (group A) in all blocks of part 2 (ß range [0.16, 0.25], all $p \leq .04$; Fig. 3A), except block 5 (ß = 0.14, $p = .099$). OP participants similarly

[1] One female IP participant was excluded from analyses because she informed the experimenter after completing the study that she had prosopagnosia (Damasio et al., 1982).

demonstrated a robust preference for group A in all blocks of part 2 (ß range [0.20, 0.35], all $p < .05$; Fig. 3B), except block 5 (ß = 0.28, $p = .054$) and block 6 (ß = 0.15, $p = .34$). In both conditions, this preference developed rapidly; we found a reliable preference for group A within the first block (IP: ß = 0.16, $p = .004$; OP: ß = 0.20, $p = .025$). However, unlike in Experiment 1, participants' preference for group A did not change as a function of exposure to the task; compared to block 1, we did not observe a reliable change in the likelihood of choosing Group A in blocks 2-6 (IP: ß range [-0.02, 0.09], all $p \geq .34$; OP: ß range [-0.05, 0.15], all $p \geq .20$).

We next examined how this use of reward information interacted with participants' use of social information, particularly given their robust preference for ingroup (IP) or outgroup (OP) faces at the end of part 1. In contrast to our findings in Experiment 1, IP participants exhibited a reliable preference for ingroup faces in block 1 (ß = 0.36, $p = 0.029$) and block 4 (ß = 0.36, $p = 0.04$) of part 2. Furthermore, compared to block 1, they were significantly less likely to choose ingroup faces by block 6 (ß = -0.24, $p = .03$). OP participants did not exhibit a reliable preference for either ingroup or outgroup faces at any point in part 2 (ß range [-0.10, 0.20], all $p \geq .17$), nor did their preferences change across blocks (ß range [-0.04, 0.26], all $p \geq .20$).

Experiment 1 vs. Experiment 2. A key difference between Experiments 1 and 2 was that in part 1 of Experiment 2, we provided participants with statistical information that supported the development of a robust preference for either ingroup or outgroup faces. To better understand the influence of this *a priori* preference on participants' subsequent decisions, we compared choice preferences in part 2 of Experiment 2 with the first six blocks of Experiment 1.

These six blocks presented identical information to participants in both experiments (i.e., the same faces and the same reward probabilities). Although participants in Experiment 2 were able to rapidly learn task-relevant reward contingencies, the presence of a prior ingroup preference resulted in an *initial* bias towards ingroup faces even when social information ceased to be relevant to the task at hand. Notably, this bias weakened as participants gained more exposure to this phase of the experiment, where social information was irrelevant. Indeed, participants no longer exhibited a preference for the ingroup by the final block, and the log-odds of choosing ingroup faces was significantly lower by the final block, in comparison to the first block (ß = -0.24, $p = .03$). Conversely, a prior preference for outgroup faces did not result in a similar bias towards outgroup faces.

Furthermore, comparing participants' use of reward information across the two experiments (Fig. 4), we found that participants exhibited a less robust learning rate in both conditions of Experiment 2 than in Experiment 1. We did not observe a reliable change in participants' preference for the high-reward group in either condition, between block 1 and every subsequent block, a notable departure from the pattern observed in Experiment 1. Indeed, for the IP condition, compared to block 1, we found a significantly weaker increase in the log-odds of choosing Group A in block 5 (ß = -0.52, $p = .004$) and a marginally weaker increase in block 6 (ß = -0.33, $p = .082$) in part 2 of Experiment 2 (versus the same blocks in Experiment 1). Similarly, for the OP condition, compared to block 1, we found a significantly weaker increase in the log-odds of choosing Group A in block 5 (ß = -0.42, $p = .049$) and a marginally weaker increase in block 6 (ß = -0.44, $p = .054$) in part 2 of Experiment 2 (versus the same blocks in experiment 1). These results suggest that previously learned social information may not only bias how participants subsequently use social information, but might also have a lasting influence on how subsequent reward information is learned and utilized over time.

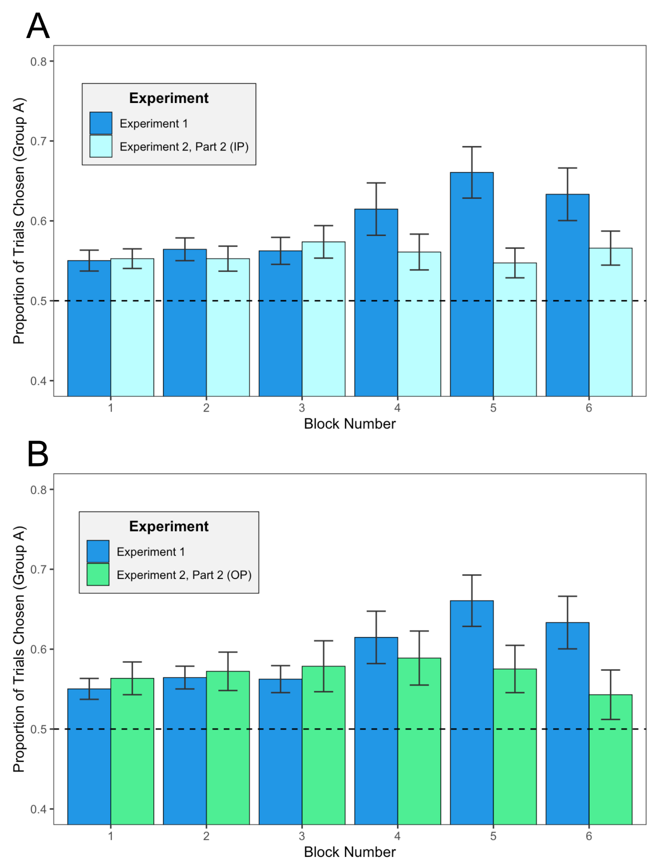

Figure 4: Comparing Experiment 1 and part 2 of Experiment 2, IP (panel A) and OP (panel B) conditions. Error bars reflect standard errors of the mean.

Discussion

We used a probabilistic decision-making task to explore how human behavior in an intergroup context is influenced by readily available but task-irrelevant social information (in the form of group membership) versus task-relevant reward information that could be learned over time. This is an important extension of previous work because real-world intergroup contexts typically encompass situations and stimuli that have been previously experienced by the observer. Our findings highlight the powerful influence of learning on participants' behavior in such contexts: Across two experiments, consistent with research in other domains (Adams et al., 2004; Hudson et al., 2007; Huttenlocher et al., 2000; Jazayeri & Shadlen, 2010; Körding & Wolpert, 2004), we found that human observers were reliably sensitive to the task-relevant environmental statistics, and they could learn and utilize this information to inform their choices. Furthermore, our findings suggest that human observers are not universally biased towards utilizing social information in all settings. Instead, when given the opportunity to learn the task-relevance of social information, people are able to utilize this learned knowledge to flexibly reweight the available sources of information.

In particular, we considered a behavior typically attributed to implicit intergroup bias—taking group membership into account when irrelevant or inappropriate (Dovidio et al., 2002)—and examined how this behavior is influenced by learning mechanisms that promote efficacy in the face of uncertainty. In most settings documenting implicit bias, participants demonstrate a bias for social categories which were relevant to previously experienced settings. Experiment 2 replicated this scenario by first establishing a preference for a social category through learning, then presenting participants with a setting in which the social categories were no longer relevant. We found that participants were able to quickly learn the relevant reward statistics in both phases of the experiment. However, Experiment 2 revealed a considerably reduced influence, in comparison to Experiment 1, of learned reward statistics when they were incongruent with previously learned group preferences. Furthermore, we found differences in participants' ability to disregard irrelevant social information depending on the social category they developed a preference for during part 1. Specifically, the presence of a prior preference for the ingroup (IP condition) resulted in an initial bias in favor of the ingroup even when social information ceased to be task-relevant (in part 2). A prior preference for outgroup faces (OP condition) did not result in a similar bias towards outgroup faces, suggesting that it was the specific preference for the ingroup (rather than a general "stickiness" of the group preferences acquired in part 1) that contributed to the observed bias in the IP condition. Notably, this IP bias was extinguished as participants gained more exposure to the task and were able to learn that social information was no longer relevant, further highlighting the role that learning can play in allowing participants to appropriately reweight available sources of information. More generally, this pattern of findings is consistent with the notion that the environmental statistics of part 1, which rendered social information diagnostic, interfered with participants' ability to subsequently extract conflicting statistical patterns in part 2. These findings also potentially provide further evidence for human observers' biased treatment of ingroup social information in intergroup contexts. In our study, environmental cues pointing to the utility of social categories influenced participants' ability to learn subsequent task-relevant (but nonsocial) information, and to appropriately weight the available sources of information, even after brief exposure in a laboratory setting. Perhaps a lifetime of exposure to societal cues regarding the relevance of social categories (Arkes & Tetlock, 2004) can be expected to have an even greater impact on observers' ability to acquire and interpret competing or conflicting nonsocial information.

Our results raise several questions for future research. The current study represents a modest step towards understanding the role that learning might play in intergroup contexts. For instance, we specifically considered a situation in which social information was both informationally and contextually irrelevant to the task at hand. In contrast, many prior studies documenting implicit bias have considered situations in which social information is informationally irrelevant but contextually relevant—for instance, where favoritism translates as a benefit that people prefer to give their ingroup (Tajfel, 1970), or where group membership can be used as a cue to infer social targets' traits, attitudes, or abilities that may be relevant to the task at hand (Stanley et al., 2011), or where the learning itself is social in nature (Spiers et al., 2017; Tajfel, 1970). Future work might use the approach described here to examine the interaction between social information and learned task-relevant information in such situations—for example, by training participants on arbitrary category distinctions within familiar social stimuli (e.g., via minimal groups; Tajfel, 1970).

Additionally, in line with the predominant approach in the literature on behavior in intergroup contexts, we chose to examine intergroup attitudes of White participants towards White versus Black targets (Amodio & Mendoza, 2010; Hewstone et al., 2002). In particular, this sample allowed us to maintain the same ingroup and outgroup categories across participants. However, an important question for future research is the extent to which our findings might be applicable to other samples. The issue of sample homogeneity is highly relevant for research on behavior in intergroup contexts because participants' responses to outgroups differ according to their current ingroup affiliation (Ray et al., 2008; Weisbuch & Ambady, 2008). Additionally, participants from different demographic backgrounds may possess different types (or strengths) of group-based expectancies about any particular group, and the makeup of these *a priori* expectancies could influence the manner in which reward and social information interact during decision-making. Finally, another productive line of work would be to examine participants' behavior using a similar experimental paradigm but including the possibility for participants to accrue losses as well as gains. Would the added possibility of accruing losses cause participants to place greater weight on readily available social information rather than learned statistics? This is relevant to high-stakes decisions in social settings, which are often associated with low-probability but high-magnitude losses (e.g., "the police officer's dilemma"; Correll et al., 2002).

Acknowledgments

We wish to thank Lantz Kilburn for his assistance with data collection.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058. https://doi.org/10.1038/nn1312

Amodio, D. M., & Mendoza, S. A. (2010). Implicit intergroup bias: Cognitive, affective, and motivational underpinnings. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications.* (pp. 353–374). The Guilford Press.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*, *15*(4), 257–278. https://doi.org/10.1207/s15327965pli1504_01

Averbeck, B. B., & Duchaine, B. (2009). Integration of Social and Utilitarian Factors in Decision Making. *Emotion*, *9*(5), 599–608. https://doi.org/10.1037/a0016509

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bernardo, J. M., & Smith, A. F. M. (2008). *Bayesian Theory*. John Wiley & Sons. https://doi.org/10.1002/9780470316870

Blascovich, J., Wyer, N. A., Swart, L. A., & Kibler, J. L. (1997). Racism and racial categorization. *Journal of Personality and Social Psychology*, *72*(6), 1364–1372. https://doi.org/10.1037/0022-3514.72.6.1364

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science*, *239*(4843), 985–992. https://doi.org/10.1126/science.239.4843.985

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, *14*(1), 1–13. https://doi.org/10.1119/1.1990764

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331–441. https://doi.org/10.1212/wnl.32.4.331

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2), 143–169. https://doi.org/10.1023/B:SORE.0000027407.70241.15

Devine, P. G. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology*, *8*(2), 88–102. https://doi.org/10.1037/1099-9809.8.2.88

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, *9*, 15. https://doi.org/10.3389/fnbeh.2015.00015

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, *23*(C), 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2007). Movement planning with probabilistic target information. *Journal of Neurophysiology*, *98*(5), 3034–3046. https://doi.org/10.1152/jn.00858.2007

Huttenlocher, J., Hedges, L. V, & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241. https://doi.org/10.1037/0096-3445.129.2.220

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (In)Accuracy in Perceptions of Groups and Individuals. *Current Directions in Psychological Science*, *24*(6), 490–497. https://doi.org/10.1177/0963721415605257

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(January), 244–247. https://doi.org/doi.org/10.1038/nature02169

Krieger, L. H. (1995). The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity. *Law Review*, 47(6), 1161–1248. https://doi.org/10.2307/1229191

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238. https://doi.org/10.1037//0033-295x.90.3.215

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience and Biobehavioral Reviews*, *35*(4), 1042–1051. https://doi.org/10.1016/j.neubiorev.2010.08.011

Nosek, B. A. (2007). Implicit-Explicit Relations. *Current Directions in Psychological Science*, *16*(2), 65–69. https://doi.org/doi.org/10.1111/j.1467-8721.2007.00477.x

Pendry, L. F., & Macrae, C. N. (1994). Stereotypes and mental life: The case of the motivated but thwarted tactician. *Journal of Experimental Social Psychology*, *30*(4), 303–325. https://doi.org/10.1006/jesp.1994.1015

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ray, D. G., Mackie, D. M., Rydell, R. J., & Smith, E. R. (2008). Changing categorization of self can change emotions about outgroups. *Journal of Experimental Social Psychology*, *44*(4), 1210–1213. https://doi.org/10.1016/j.jesp.2008.03.014

Reber, A. S. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Spencer, S., Fein, S., Wolfe, C., Fong, C., & Dunn, D. (1998). Automatic activation of stereotypes- The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*(11), 1139–1152. https://doi.org/10.1177/01461672982411001

Spiers, H. J., Love, B. C., Le Pelley, M. E., Gibb, C. E., & Murphy, R. A. (2017). Anterior temporal lobe tracks the formation of prejudice. *Journal of Cognitive Neuroscience*, *29*(3), 530–544. https://doi.org/10.1162/jocn_a_01056

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, *108*(19), 7710–7715. https://doi.org/10.1073/pnas.1014345108

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102. https://doi.org/10.1038/scientificamerican1170-96

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face Preference at Birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 892–903. https://doi.org/10.1037/0096-1523.22.4.892

Weisbuch, M., & Ambady, N. (2008). Affective Divergence: Automatic Responses to Others' Emotions Depend on Group Membership. *Journal of Personality and Social Psychology*, *95*(5), 1063–1079. https://doi.org/10.1037/a0011993

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial Bias in Judgments of Physical Size and Formidability: From Size to Threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. https://doi.org/10.1037/pspi0000092.supp

Zebrowitz, L. A., & Montepare, J. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kendrick (Eds.), *Evolution and Social Psychology* (pp. 81–113). Psychosocial Press.