American Psychological Association
Journal of Experimental Psychology: General

0096-3445/19/$12.00
2019, Vol. 148, No. 5, 845–862
http://dx.doi.org/10.1037/xge0000601

The Group-Motivated Sampler

Robin Bergh
Uppsala University and Harvard University

Marcus Lindskog
Uppsala University

Does ingroup favoritism reflect experience or some preset motivation? The latter possibility is well examined in social psychology, but models from cognitive psychology suggest that unrepresentative samples of experience can generate biases even in the absence of motivational concerns. It remains unclear, however, how motivation and initially sampled experiences interact when both influences are possible, and people encounter new groups. Extending classic arguments about motivated information gathering, we propose that people can be described as "group-motivated samplers"—marked by a tendency to primarily seek out information about one's own group, and to attend more to information that portrays the ingroup in a positive light. Four experiments showed that information seeking almost always starts with the ingroup, and that people chose to gather more information from the ingroup compared to an outgroup. In subsequent group evaluations, people were excessively positive about ingroups giving a good initial impression. Participants were also fairly accurate, on average, about the direction and magnitude of group differences when the ingroup was de facto better, but downplayed those differences in the opposite situation. Further analyses indicated that first experiences led to biased evaluations because people failed to discount for nonrepresentative (positive) ingroup experiences, whereas interpretive biases seem responsible for evaluations based on belonging to a better/worse performing group. Taken together, while social psychologists know that people tend to portray ingroups in a flattering light, we show how people selectively incorporate early experiences to build those impressions.

Keywords: ingroup biases, motivated cognition, experience sampling

Imagine starting a job in a new field and you first meet people from your new workplace at a conference, along with many others in the same profession. Or imagine finding out that most of your prehistoric ancestors were Yamnaya. With little knowledge about this group, you might explore online forums for ancestry testing to learn how people with this background differ from others. In both scenarios, what do you think your impressions would be after the initial encounters with members of these new groups?

Addressing that question, we examine how ingroup biases emerge through interactions between initially sampled experiences and preexisting motivations. This follows naturally from the classic notion that social learning is regulated by motivations and beliefs (e.g., Higgins & Bargh, 1987; Snyder & Swann, 1978). Yet, there is little research on how this plays out in novel group categorizations. A few studies have examined what people recall from encounters with members from novel groups, and how these experiences are associated with ingroup favoritism (e.g., Gramzow, Sedikides, & Gaertner, 2001; Howard & Rothbart, 1980; Schaller & Maass, 1989). These studies have not involved any active information seeking, however, and instead reflect what people learn from a fixed set of stimuli. In the example of learning about groups of people at a conference, this is analogous to studying individuals who have the exact same interpersonal encounters. Such a setup makes for neat experiments, but it leaves the following questions unanswered: What happens when people have a choice of whom to meet? Whom will they start to socialize with and what groups will they meet the most people from? What are the consequences of starting to socialize with new people from one's own organization, and socializing with more people from that group?

In this article, we examined biases that emerge in the process of gathering information about groups, and in doing so we integrate two theoretical and methodological perspectives marked by little cross-talk to date. One of these perspectives emphasizes motivation: Ingroup biases arise because people are intrinsically more interested in their own group, because people derive psychological or tangible benefits from their membership in a group, and because people interpret events based on what they want believe about groups (e.g., Brewer, 1979; Hastorf & Cantril, 1954; Tajfel & Turner, 1979; Yamagishi & Mifune, 2009). From this perspective,

Robin Bergh, Department of Psychology, Uppsala University, and Department of Psychology, Harvard University; Marcus Lindskog, Department of Psychology, Uppsala University.

Robin Bergh and Marcus Lindskog contributed equally to this work. Robin Bergh and Marcus Lindskog developed the study concept and designed the studies. Marcus Lindskog supervised the lab data collection, Robin Bergh collected the online data. Robin Bergh and Marcus Lindskog performed the data analyses and wrote the manuscript. We thank Marie Lundström, Elina Stengård, Sofia Vesterqvist, and Johanna Wångenberg for help with the data collection in the lab experiments and Spencer Stone for programming assistance for the online sampling tasks. We are also thankful to Mirta Galesic for valuable comments on earlier versions of the article. Preliminary findings of this work were presented at the annual meeting of the Society of Experimental Social Psychology.

Correspondence concerning this article should be addressed to Marcus Lindskog, Department of Psychology, Uppsala University, P.O. Box 1225, Uppsala SE-751 42, Sweden. E-mail: marcus.lindskog@psyk.uu.se

ingroup biases can even emerge without any personal experience with ingroup or outgroup members (e.g., Tajfel, Billig, Bundy, & Flament, 1971). From the second perspective, ingroup favoritism, like a number of other cognitive biases, emerges because of nonmotivational inferential errors based on nonrepresentative data (e.g., Denrell, 2005; Fiedler, 2000). From this perspective, ingroup biases can emerge without motivation. This does not imply that the two perspectives are mutually exclusive, however. They can also be seen as independent and complementary (as acknowledged by proponents of each perspective; see Denrell, 2005; Dunham, Baron, & Carey, 2011). Furthermore, motivational factors can be viewed as moderators in the process of drawing inferences from one's experiences (Gramzow et al., 2001; Howard & Rothbart, 1980). Our goal here is to also consider another kind of motivational interaction, occurring earlier on in the inferential process. In particular, we stress that there are (motivated) *choices* that people make when gathering information about groups, which can also bias evaluations. Next, we describe the group-motivational and experiential perspectives in greater detail, followed by a section on motivated information seeking.

Group-Motivational Perspectives

Classic minimal group research has examined biases when participants have no previous experience with the groups or their members (as individuals are anonymous and the groups are novel and arbitrarily created; Brewer, 1979; Tajfel et al., 1971). Thus, there is no information to feed a data-driven process, and any biases should therefore reflect motivational processing (although the precise mechanisms and motives have been debated; Cadinu & Rothbart, 1996; Yamagishi & Mifune, 2009). The developmental literature also suggests that young children show a mental preparedness for ingroup favoritism—biases do not seem to require prolonged learning (e.g., Dunham, Baron, & Banaji, 2008; Dunham et al., 2011). A default desire to view the ingroup positively is also evident in categorizations of individuals whose group membership is ambiguous, and in evaluations of new ingroup and outgroup members. For example, certain individuals are seen as part of the ingroup when they do good things (e.g., winning Olympic medals), but viewed as outgroup members when doing bad things (e.g., cheating or killing civilians; see, e.g., Kteily, Cotterill, Sidanius, Sheehy-Skeffington, & Bergh, 2014; Stelzl, Janes, & Seligman, 2008). When group membership is nonambiguous, it has been shown that people ignore negative information when they learn about new ingroup members, but not when they learn about new outgroup members (Hughes, Zaki, & Ambady, 2017).

This research fits within a long tradition of studying motivated cognition in social psychology (e.g., Fiske, 1993b; Fiske & Neuberg, 1990), in which stereotypes and prejudices are examined in a top-down fashion, as a function of preexisting goals and beliefs (e.g., Kunda & Oleson, 1995; Tajfel & Turner, 1979). Social–cognitive models of individual impression formation, that explicitly acknowledge data-driven processes, tend to also describe the role of category representations and group motives as exogenous to the analysis (Brewer, 1988; Fiske & Neuberg, 1990). In other words, impression formation does not start with an agnostic search for information, but rather with what a person already thinks s/he knows and what she wants to believe about social reality—for instance, that one's own group is better than other groups. Even without an explicit motive for ingroups to trump outgroups, biases can also emerge from more innocuous motivations related to "ingroup love" (Brewer, 1999), by which ingroups may simply receive more attention.

Sampling Perspectives

In cognitive psychology, there are several formal models describing how people learn about new categories through experience, especially in nonsocial situations (for a review see Ashby & Maddox, 2005). Some of these focus on classification, and learning to distinguish between members of different categories (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Fazio, Eiser, & Shook, 2004; Juslin, Karlsson, & Olsson, 2008; Juslin, Olsson, & Olsson, 2003). Another class of models deal with situations in which the existence of categories is known (e.g., different organizations present at a conference), and people make inferences about them via experience with individual members (Denrell, 2005; Fiedler, 1996; Kalish, Rogers, Lang, & Zhu, 2011). The latter are often referred to as sampling models.

Sampling refers the process of selecting a subset of stimuli from a universe of available ones. Fiedler and Juslin (2006) argued that understanding which stimuli are selected in a given situation, is crucial for understanding human behavior. By now, a large literature has used the insight that people experience only a subset of all available stimuli to explain a range of behaviors. These findings include, but are not limited to, how overconfidence may emerge from sampling information from memory (Juslin, Winman, & Hansson, 2007), how deciding between lotteries is influenced by self-sampling of outcomes (Hertwig & Erev, 2009), and how variability comes to be underestimated when people do not take the properties of samples into account (Kareev, Arnon, & Horwitz-Zeliger, 2002). Further, a growing literature has shown how experiencing only a subset of all stimuli may come to influence the perception of groups (Fiedler, 2000; Denrell, 2005; Le Mens & Denrell, 2011).

In contrast to the group-motivational perspective outlined earlier, Fiedler (1996) argued that biased judgments, for instance about groups, may arise because people are unaware of the lack of representativeness in their sampled experiences, and because they fail to correct for this deficit (see also Fiedler, 1996; Fiedler & Juslin, 2006; Juslin et al., 2007; Lindskog, Winman, & Juslin, 2013; Tversky & Kahneman, 1971). That is, people can be considered as being naïve intuitive statisticians when learning from new data (Fiedler & Juslin, 2006; Fiedler, 2000; Juslin et al., 2007). The metaphor suggests that although people can veridically represent the data they encounter, they are naïve with respect to the processes that generate it, which can give rise to biases.

Similarly, Denrell (2005) proposed a model in which biased beliefs stem from people sampling more instances from their own group than other groups. The logic of his model is that people will not continue interacting with someone whom they have a negative initial impression of. Thus, negative initial impressions, that may be wrong, are unlikely to be corrected unless "incidental" encounters happen to take place, beyond one's personal choice. As people are more likely to interact with people from their own group, initial negative impressions of ingroup members should be more easily corrected than those of outgroup members (see also Le Mens &

Denrell, 2011). The common theme in the sampling perspective is that evaluative biases can arise in a data-driven fashion. More broadly, sampling models are growing in popularity to explain many social judgment biases, thereby providing alternative (or rather complementary) accounts to the classic motivational assumptions in the literature (e.g., Dawtry, Sutton, & Sibley, 2015; Galesic, Olsson, & Rieskamp, 2012, 2013; Kutzner & Fiedler, 2016; Pachur, Hertwig, & Rieskamp, 2013).

Motivated Sampling

According to the sampling models discussed above, biased impression about groups can arise without people making an active choice about which information to search for, but rather as a result of how information is distributed in the environment (e.g., Fiedler, 1996, 2000) or how information search is terminated (e.g., Denrell, 2005). People are not necessarily passive, however, in their search for information. Sometimes their sampling is motivated by an explicit or implicit information search strategy (Klayman & Ha, 1987; Oaksford & Chater, 1994). For example, it has been suggested that people are more inclined to seek confirming than disconfirming evidence in a range of situations, an inclination often referred to as confirmation bias. Klayman and Ha (1987), however, showed that a wide range of such findings could be explained by people using a general positive test strategy, rather than seeking confirming evidence per se. According to this account, people test a hypothesis about the world by sampling cases that are expected to have a certain property, rather than cases that are expected not to. For example, to test the hypothesis that those with Yamnaya ancestry are kind people, you would probably try to find someone who is Yamnaya and check if she or he is kind, rather than find someone who is not kind to make sure she or he is not Yamnaya.

More generally, this problem concerns testing if a rule of the form "If *p*, then *q*" is correct or not. Starting with Wason's (1960) selection task a long line of research has investigated if people use a rational strategy when testing such a rule. The conclusion from many studies is that people in general do not solve the task in a completely rational way. More specifically, they do not sufficiently seek out evidence that can falsify the rule. In terms of a positive test strategy, this means not sufficiently testing cases that are not expected to have the property *q*. The analysis by Klayman and Ha (1987), however, indicated that although a positive test strategy can sometimes lead to biases, it is a surprisingly good heuristic in many situations to determine if a hypothesis is true or false. Similarly, Oaksford and Chater (1994) showed that a wide range of results in Wason's selection task could be accounted for by a cognitive model that assumes people are sampling information that gives the highest information gain, rather than by a systematic bias.

These findings suggest that in several situations people will probably, for various good reasons, be motivated to selectively sample information when testing a hypothesis—for example, about groups. This does not guarantee desirable conclusions (e.g., that the ingroup is better), but it makes such beliefs harder to disconfirm. More broadly, previous research has acknowledged the influence of both motivational and data-driven processes in general (e.g., Fiedler, 1996), but there have been few attempts to date to integrate them when studying group biases.

The Current Studies

In this article, we integrate motivational and sampling perspectives on ingroup biases by examining how people gather information about novel groups. Thus, rather than examining what people do in the absence of information (a classic premise in intergroup research; e.g., Tajfel et al., 1971), or in the absence of social motivation (a common premise from a sampling perspective), we examine how people gather and interpret information as "group-motivated samplers." This notion signifies an active search for information about groups, and how it is likely skewed by differential interest in ingroups and outgroups. The notion of a group-motivated sampler does not deny the explanatory power of naïve and nonintentional errors in the inferences of limited personal experience (Fiedler, 2000), but we predict that people end up with nonrepresentative samples in part because of a greater intrinsic interest (i.e., motivation) in their own group. Thus, aside from motivations having direct effects on group evaluations (Tajfel & Turner, 1979), as well as distorting the recall of ingroup and outgroup information (e.g., Howard & Rothbart, 1980), we suggest it also drives what experiences people have in the first place.

More broadly, the motivations that are commonly proposed to underpin group biases are fairly blunt and overtly self-serving. The basic premise is that people explicitly *want* their group to be or do better, because that is also better for oneself (e.g., giving the actor better self-esteem and/or more resources; Sherif, 1966; Tajfel & Turner, 1979; Yamagishi & Mifune, 2009). The motivation we propose here is subtler: A simple, and potentially implicit, desire to learn more about ingroups might be enough to generate evaluative group biases. This "innocuous motivation" viewpoint puts the current article on a middle ground between theories of prejudice that emphasize blunt motivations (e.g., Sidanius & Pratto, 1999; Tajfel & Turner, 1979) and those that describe it as largely accidental (e.g., Denrell, 2005; Fiedler, 2000). Furthermore, the current studies extend what has been previously investigated in two fields of research with little cross-talk to date. From the perspective of the motivational literature, a novel aspect of this research is the incorporation of sampling models to study information acquisition. In relation to that literature, another important feature of these studies is that we examine what people learn from encounters with multiple group members, rather than a single individual from ingroup or outgroup (see, e.g., Hughes et al., 2017). From the perspective of the judgment and decision-making literature, a novel feature is the incorporation of the minimal group paradigm to study group-based motivation processes.

With these premises in mind, we conducted four experiments. In the first two experiments, participants sampled information about how members of two groups had rated a restaurant, and we examined if and how they formed impressions that one group (their own) provided more helpful advice. In the last two experiments, participants sampled information that is more consequential. In the third experiment, the information constituted performances on a weapon-screening task, implying that low scores should also carry more weight than in the first two studies. In the fourth, participants sampled information on a variable with clear parallels to real-life stereotyping, namely intelligence (e.g., Devine & Elliot, 1995). In all experiments, participants were free to sample as many (or few) pieces of information as they wanted from each of the two groups. As such, the sampling in our experiments mimics the way people often search information in most real life situation, where they often neither have the time nor the cognitive capacity to seek complete information.

Two outcomes in each experiment are relevant for our analysis: The sampling behavior in itself, and subsequent group evaluations. For the first outcome, we hypothesized that most participants would (a) start to sample from their own group and (b) seek out more information from the ingroup. This is to suggest that people are *by default* more interested in their own group (Brewer, 1999), in line with an innocuous motivated cognition account. So while ingroup favoritism could in theory be entirely limited to data-driven biases (as supported by simulated data; Denrell, 2005; Le Mens & Denrell, 2011), we suggest that personal choices introduce a bias is present from the very first sample and, in practice, drive sampling behaviors throughout the information seeking process.

To examine how biased sampling may lead to biased group evaluations, we manipulated if the first piece of information collected would be positive or negative. To the extent that people start sampling from their own group, this should influence their overall (average) experience with the ingroup, but not the outgroup. Unless people choose to gather a lot of information, and nonsocial sampling research suggests that is uncommon (e.g., Hertwig & Erev, 2009), this introduces a systematic bias in people's experiences with the ingroup. To be more specific, if a person initially receives overly positive information about the ingroup, and only samples a few pieces of information, then the person is likely to overestimate how good the ingroup is—even in the absence of any interpretive bias. In contrast, a person who first has an overly negative experience should be equally likely to underestimate the ingroup, in the absence of an inferential bias. We predicted, however, that people will attend more to positive than negative ingroup information, and hence report that the ingroup is better when the first sample is positive, but not report the ingroup as worse when the first sample is negative. From a sampling perspective, we would not expect differences in the outgroup evaluations, simply because initial sampling from the ingroup should produce experiential differences in the ingroup only. This prediction reconciles the emphasis on data-driven judgments in the sampling literature with motivated biases, both at the information gathering and inferential step in the evaluation process.

The notion that first positive samples lead to ingroup biases, while negative ones do not, is based on both a sampling bias and an interpretative bias. Such a result would be difficult to explain by either perspective in isolation. Again, if people only display sampling biases, but no interpretive bias, then they should be equally likely to over- and underestimate the ingroup. In contrast, if people only show an interpretive bias, but no sampling bias, then we would expect the first sample to be randomly selected from the ingroup versus outgroup. In this scenario, initial negative outgroup information should also have a similar effect as positive ingroup information. This is because observers would not know (unless they sample a fair amount) that their initial experiences are positively skewed for the ingroup versus negatively skewed for the outgroup. In both cases, they would just see that the ingroup is better, and if this is the kind of information they are attuned to, they should all show a similar evaluative bias. In contrast, if evaluative biases are only observed for positive initial information then sampling biases should be responsible for that.

Another empirical implication is that both ingroup and outgroup experiences should mediate the experimental effects for unbiased samplers, while only ingroup experiences should mediate the effects for biased samplers (see the section The Role of Experience—Auxiliary Analyses). Taken together, interpretive biases feed on particular experiences and we can make divergent predictions what those experiences would be when people are biased (only ingroup experiences generate bias) or unbiased samplers (ingroup and outgroup experiences can both generate bias). This makes the current inquiry different from previous social psychological studies that focus entirely on interpretive biases (see, e.g., Howard & Rothbart, 1980; Schaller & Maass, 1989).

In Experiment 1, participants sampled information from two identical group distributions. In Experiments 2 to 4, we added another manipulation of the sampled information: whether the ingroup was better or worse overall. The bias introduced by the first experience should get weaker the more information people gather, but true mean differences between groups would not reveal such a pattern. Nevertheless, we could make a similar prediction about attending or not to positive and negative information about the ingroup. We hypothesized that those in the better-ingroup condition would detect and acknowledge the difference, whereas those in the worse-ingroup condition would downplay it.

Taken together, both initially positive/negative information about the ingroup, as well as a better/worse ingroup overall, represents information that aligns or misaligns with a priori beliefs and motivations. We propose that people selectively seek out, and attend to information that affirms a positive picture of the ingroup (in line with a positive test strategy, Klayman & Ha, 1987). Thus, we argue that a group-motivated sampler is capable of data-driven inferences about groups, although his or her experiences will be nonrepresentative because of a greater interest in the ingroup. We further argue that the capacity for accurate inferences based on one's experience is employed differently as function of what the person believes a priori, or wants to believe. Put differently, in one sense the group-motivated sampler is like a naïve intuitive statistician (Fiedler, 2000; Fiedler & Juslin, 2006; Juslin et al., 2007) – drawing logical inferences about a population based on samples, while being naïve to how one's sample estimates deviate from the population ones. The motivated aspect is twofold. First, the sampling biases are systematic due to preferences, and not random. Second, the inferences are more likely to be wrong in one direction (favoring the ingroup), than the other (favoring the outgroup).

Experiment 1

Method

Participants. Participants ($N = 40$) were mainly (nonpsychology) students at a large Scandinavian university. They received a movie voucher (~€10) for their participation. The sample had a mean age of 26.1 years ($SD = 9.9$) and consisted of 60% females. Other studies in Scandinavia indicate strong and robust evaluative biases in minimal groups ($\eta_p > .15$; Bergh, Akrami, Sidanius, & Sibley, 2016).

Participants were recruited by flyers posted on notice boards at various departments and at places accessible to the general public, such as the public library. Before recruitment, participants were screened for studies in psychology, or participation in studies similar to the current one. Only participants without such experience were invited to participate.

Procedure and materials. The experiment was programmed in Python using the *Expyriment* framework (Krause & Lindemann, 2014). The experiment consisted of four parts, a minimal group manipulation, sampling, judgment, as well as personality and demographics. All parts were completed individually.

During the minimal group manipulation, participants were first informed that they were going to take a newly developed psychological test named "GHP-type indicator." Participants were told that this test could divide people into two broad categories, GHP-J and GHP-P, based on their preference for pieces of art (for a similar methodology, see also Tajfel et al., 1971). No further information was given about the test or the categories. After these instructions participants first rated their liking of six pieces of art on a Likert scale, and they were also given three forced choice items comparing pairs of art pieces. Participants were subsequently given feedback on category membership, based on (bogus) art preferences. Half of the participants were randomly assigned to the GHP-P category and the other half to the GHP-J category.

In the sampling task, participants were asked to gather information about a restaurant ("The Knife and the Fork") to form an opinion about the quality of the restaurant. They were told that the information about the restaurant came from a database of responses from previous studies and participants who had also been categorized as belonging to either GHP-J or GHP-P. Participants were informed that they would see two buttons with the labels "GHP-J" and "GHP-P" and that pressing one of these buttons would present them with two pieces of information from a previous participant in the corresponding GHP-category. The two pieces of information was a rating of the restaurant and a helpfulness score for the previous participant. The helpfulness score was explained to indicate the trustworthiness of the previous participant (analogous to variable reputations of different contributors in online forums). After having been shown an example rating, participants could sample freely from the two categories until they felt they had collected a sufficient amount of information to make a judgment about the restaurant and previous raters. The position of the buttons labeled GHP-J and GHP-P was counterbalanced across participants. An illustration of the information presented during sampling is shown in Figure 1.

Ratings and helpfulness scores for the two categories were sets of 120 randomly drawn numbers in the range of 1 to 5. Ratings were constrained to be integers while helpfulness scores were given with one decimal. The sets were approximately normally distributed with a mean of 3 and standard deviation of 1. The presentation order of the values was individually randomized for each participant. To evaluate the impact of the first piece of information encountered during the sampling phase, we manipulated the first helpfulness rating between-subjects to be either positive (4.5) or negative (1.5).

In the next phase of the experiment, participants were asked to make a series of judgments about the information acquired during sampling. They were asked to make point estimates about the quality of the restaurant, the average rating of the restaurant, and the average helpfulness of the previous participants from each group (i.e., overall impression of GHP-J and GHP-P). The order of judgments was counterbalanced over participants.

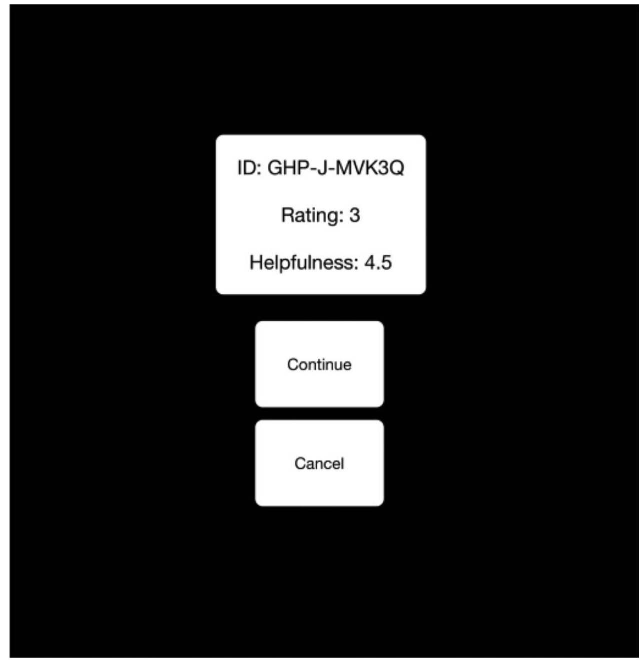

Figure 1. Illustration of how the information about rating and helpfulness was presented to participants during sampling.

In the last part of the experiment participants answered a few personality questions[1] and a series of demographic questions, including two items to assess group identification ("How strongly do you identify with your GHP-type?" and "How meaningful do you think that GHP-types are for describing different people?"; $\alpha = .67$), along with two items to assess gender identification. The latter was done to disguise our specific interest for the minimal groups. Participants were also prompted to recall their assigned GHP-type, and everyone did so correctly. We finished by asking if participants had any familiarity with the groups before the onset of the experiment. Answers were expected to be no, alternative some vague (imagined) recollection, and everyone responded accordingly. After the experimental procedure, participants were shown a text on screen that explained the true intent of the experiment. They could also ask any additional questions to the experimenter. The experiment took approximately 20 minutes to complete. Experiment materials, deidentified raw data, and analysis scripts for all four experiments are publicly available for download at https://osf.io/c5qk4/.

Results

Our first hypothesis was that most participants would start sampling from their own group. In line with this hypothesis, 85% of participants chose to sample the first piece of information from

[1] Here, as in Experiment 2, we included a measure of openness to experience (see Lee & Ashton, 2004) to examine systematic individual differences in sampling behaviors and/or the group evaluations. We suspected that open-minded individuals might have weaker sampling biases (by virtue of sampling more from the outgroup), but we found no evidence thereof.

their own group, binomial test $H_0 = .5$: $p < .001$, 95% confidence interval (CI) [.70, .94]. With respect to the amount of information participants sampled, we expected participants to seek more information from the ingroup, with the possibility that this effect could be influenced by whether the first piece of information participants was shown was either helpful (4.5) or not (1.5). On average participants drew 25.0 samples, $SD = 39.5$. Of these, 15.8, $SD = 25.3$, were sampled from the ingroup and 9.2, $SD = 18.1$, from the outgroup.

We formally examined the impact of group membership on drawn samples using a negative binomial generalized linear mixed-effects model with valence of first sample (positive [helpful rating] vs. negative [not helpful rating]), and sampled group (ingroup vs. outgroup) as fixed factors and subject as random factor. We used such a model to account for the zero-inflated and overdispersed count distribution. The model was estimated using the *glmer.nb* function of the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). This analysis revealed a main effect of Sampled group, reaffirming that participants sampled significantly more from the ingroup than the outgroup, $\chi^2(1) = 8.2$, $p = .004$, $b = -.22$, $SE = .08$, $z = 2.88$. All other main and interaction effects were nonsignificant, $ps > .33$.[2]

After the sampling process, participants were asked to evaluate the helpfulness of the two groups.[3] Recall that the actual expected value of the helpfulness score was the same in both groups ($M = 3.0$). We investigated the extent to which participants were biased in their estimates of helpfulness by submitting the rated helpfulness for the ingroup and outgroup to two separate single sample *t* tests ($H_0 = 3$). This analysis revealed no bias in judgments for either the ingroup, $M = 3.2$, $t(38) = 1.3$, $p = .19$, nor the outgroup, $M = 3.0$, $t(38) = .05$, $p = .96$, at least not aggregated across positive and negative initial experiences. Next we examined the possible effect of valence of the first sample, and an interaction with evaluated group on these evaluations. We tested these effects using a generalized linear mixed-effects model (fitted with the *lmer* function of the *lme4* package in R) with valence of first sample (positive [helpful rating] vs. negative [unhelpful rating]) and evaluated group (ingroup vs. outgroup) as fixed factors and subject as random factor. This analysis showed no significant main effects (both $ps > .11$). However, the interaction between the valence of first sample and group was significant, $\chi^2(1) = 6.89$, $p = .009$, $b = .18$, $SE = .07$, $t = 2.67$). As is evident in Figure 2A, there was no difference in judgments of helpfulness if the first sample was negative, $t(18) = 0.88$, $p = .39$. However, when the first sample was positive participants rated the helpfulness of the ingroup as higher than that of the outgroup (and higher than the actual helpfulness of the ingroup), *M*s 3.50 and 2.98; $t(19) = 3.32$, $p = .004$.

Discussion

These results suggest that the vast majority of people start gathering information from their own group when facing a novel categorization, in line with our first motivational hypothesis. Further, participants chose to gather more information in total from their own group, as compared with the outgroup. In terms of final group evaluations, initial positive ingroup information led to biased evaluations (excess positivity about the ingroup), whereas initial negative information produced no difference. This is in line with the notion of positive testing, and in particular the notion that people pay more attention to information that affirms what they want to believe—that the ingroup is better.

Aside from the observation that positive sampling led to excessive ingroup optimism, most evaluations were close to the true values and seemingly data-driven. However, such responses are also expected from a person giving a socially desirable response (expressing that both groups are moderately [and equally] helpful). Still, given that people are typically willing to reveal ingroup favoritism in minimal group situations (see Mullen, Brown, & Smith, 1992), we doubt that social desirability would be of great concern here. Perhaps a bigger concern is that respondents may have taken little note of the sampled data and just guessed about the group characteristics at the midpoint of the response scale. What speaks against such guessing is that those who received a positive first sample deviated from the midpoint in the estimates of ingroup helpfulness. Nevertheless, in Experiment 2 we introduced an underlying actual group difference to rule out both of these alternative explanations. This also served to illustrate another motivational force in how people interpret the data they have just sampled.

Experiment 2

This experiment was identical to Experiment 1, except we increased the sample size and made one group objectively better (more helpful) than the other. We hypothesized that participants are capable of being fairly accurate about the data they sample, but only show this capacity when it fits an a priori assumption that one's own group is better. In other words, we predicted that those belonging to the better group would pick up on the underlying difference, whereas those in worse group would downplay the same difference in their responses. In extension, we expected the better group to be more accurate than the worse group, in terms of the magnitude of the group difference.

This would represent another example of an interaction between motivational processes and experiental learning, such that people form group impressions selectively based on data they want to observe. Importantly, such an interaction would not be expected if people were "data blind" and just guessed at the midpoint of the scale, or if they just strived to appear unbiased (in both case, the ingroup should be rated the same as the outgroup, regardless of what group that was actually better). Our hypothesis about differential accuracy when the ingroup is better or worse would also stand in opposition to the possibility that ingroup biases are strictly data-driven (in which case members of both groups should be equally likely to identify the better group; see also Denrell, 2005).

[2] Analyzing the effect of sampled group on number of drawn samples separately for the J and P conditions instead, showed that while members of J sampled significantly more from the J category than the P category, $\chi^2(1) = 8.22$, $p = .004$, $b = -.24$, $SE = .08$, $z = 3.1$, the reverse was true for members of P, $\chi^2(1) = 5.11$, $p = .02$, $b = .31$, $SE = .13$, $z = 2.3$, indicating a cross-over interaction.

[3] Because the main interest in the current study was the impression participants formed about other participants, rather than about the restaurant per se, our main variable of interest was the evaluation of helpfulness. Accordingly, we considered the evaluation of the restaurant to be primarily filler information to the cover story and no analyses were carried out on those ratings.

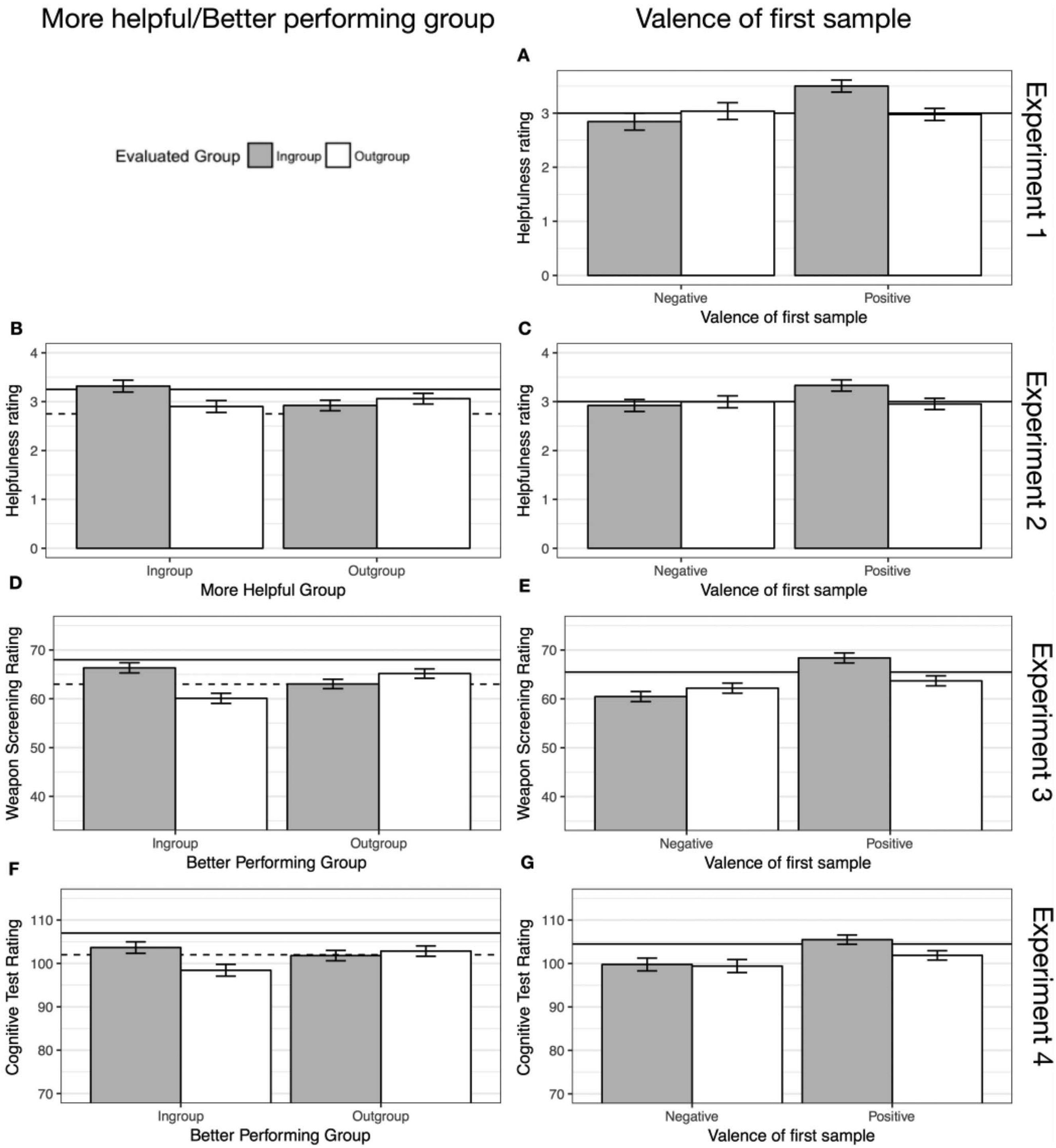

Figure 2. Evaluated performance of other participants as a function of evaluated group and valence of the first sample (Panels A, C, E, and G) or more helpful/better performing group (Panels B, D, and F) in Experiment 1 (Panel A), Experiment 2 (Panel B and C), Experiment 3 (Panel D and E), and Experiment 4 (Panel F and G). Vertical bars denote standard error of the mean. Horizontal lines in Panels B, D, and F denote expected value of actual score for the more helpful/better performing (solid) and less helpful/worse performing (dashed) groups, respectively, whereas solid lines in Panels A, C, E, and G denote the expected value in the entire sample, that is, the grand mean in the sampled information.

Finally, we expected to replicate that ingroups receive overly optimistic evaluations when the first sample is positive.

Method

Participants. Participants ($N = 63$) were mainly (nonpsychology) students. They received a movie voucher (~€10) for their participation. The sample had a mean age of 24.5 years ($SD = 6.7$) and consisted of 73% female participants. Participants were recruited in the same way as for Experiment 1.

We did not perform an a priori power analysis for this experiment because (a) Experiment 1 indicated good observed power for all the hypothesized effects there (>.74), and (b) we had no educated guess about how strong the effect of the new manipulation would be. Instead of an arbitrary guess about the new effect size, we made the equally arbitrary decision to increase the sample size by roughly 50%. The obtained sample size maintains an equivalent power for the previous Valence × Sampled Group interaction in Experiment 1, given the more complex design.

Procedure and materials. The procedure and materials for Experiment 2 was the same as in Experiment 1 with one important change. First, in Experiment 1 both GHP categories had the same average helpfulness ($M = 3$, $SD = 1$). In Experiment 2 this was changed such that one category was more helpful ($M = 3.25$, $SD = 1$) than the other ($M = 2.75$, $SD = 1$). We manipulated which category was more helpful between-subjects, such that for half of the participants their own category was more helpful, while for the other half the other category was more helpful. For example, for a participant in the own better condition who had been assigned to the GHP-J category, the average helpfulness score of previous GHP-J and GHP-P participants were 3.25 and 2.75, respectively.

Results

We first investigated which group participants began to gather information from. Replicating the finding from Experiment 1, a very large proportion of participants, 92%, picked the first sample from their ingroup, binomial test $H_0 = .5$: $p < .001$, 95% CI [.82, .97]. As in Experiment 1, we further expected participants to seek more information from their own group. We investigated this possibility and the effects of valence of the first sample and more helpful group on sampling behavior by fitting a negative binomial generalized linear mixed-effects model with valence of first sample (positive [helpful rating] vs. negative [not helpful rating]), more helpful group (ingroup vs. outgroup), and sampled group (ingroup vs. outgroup) as fixed factors and subject as random factor. We found one significant effect, the main effect of sampled group, $\chi^2(1) = 14.2$, $p < .001$, $b = -.22$, $SE = .055$, $z = 4.1$; all other ps > .11. Participants sampled significantly more from the ingroup, $M = 16.1$, $SD = 21.4$, than the outgroup, $M = 8.9$, $SD = 9.3$.

As in Experiment 1, participants were asked to evaluate the helpfulness of the two groups after their choice to end the sampling. We tested possible effects of more helpful group and valence of the first sample on these judgments by means of a generalized linear mixed model with valence of first sample (positive [helpful rating] vs. negative [not helpful rating]), more helpful group (ingroup vs. outgroup), and evaluated group (ingroup vs. outgroup) as fixed factors and subject as random factor. This analysis showed two significant effects; all other ps > .18. First, there was an interaction between more helpful group and evaluated group; $\chi^2(1) = 5.58$, $p = .018$, $b = .14$, $SE = .06$, $t = 2.45$. The interaction is illustrated in Figure 2B. Participants in the better (more helpful) ingroup condition were fairly accurate about the direction and magnitude of the difference, and hence evaluated the ingroup as more helpful, $t(32) = 2.43$, $p = .02$ (Ms = 3.31 and 2.90). In contrast, participants in the less helpful ingroup condition showed no difference in their ratings of helpfulness, Ms = 2.92 and 3.01, $t(27) = 0.91$, $p = .37$.

It is possible that this pattern of results could occur because of some artifact related to the fact that participants sampled more from the ingroup than the outgroup. We investigated this possibility by simulating 10,000 experiments where participants exhibited the same sampling behavior as in Experiment 2, but where unbiased and reported the average helpfulness of their observations (code for the simulations are available for download at https://osf.io/c5qk4/). These simulations showed that the same pattern of results, a difference when the ingroup is more helpful but not when the outgroup is, was found in only 2% of the simulated experiments. It is thus unlikely that our results emerged due to differences in the amount of sampling from the two groups, if responses were entirely data-driven (as opposed to also influenced by motivational processes).

In addition to the significant More Helpful Group × Evaluated Group interaction, there was a significant, $\chi^2(1) = 4.24$, $p = .04$, $b = .11$, $SE = .06$, $t = 2.00$, interaction between valence of the first sample and evaluated group (see Figure 2C). As in Experiment 1, this interaction reflected that participants in the positive valence condition evaluated the ingroup as being more helpful than the outgroup, Ms = 3.34 and 2.95, respectively, $t(31) = 2.32$, $p = .03$, while no such difference was present in the negative valence condition, Ms = 2.92 and 3.00, respectively, $t(28) = 0.43$, $p = .66$.

Finally, we examined how participants' evaluations might deviate from what the data would suggest in their individual (random) set of experiences. In other words, while the previous analyses indicate biases in relation to the aggregate expected values, one could also ask if (and in what direction) individuals are biased, relative to what they have personally seen—a data-driven respondent should in many cases deviate from the true group means. To further explore this issue we calculated, for each participant, the signed difference between the estimated helpfulness and the individually observed helpfulness for the ingroup and outgroup, respectively. These deviation scores were submitted to the same analysis as the original ratings of helpfulness.[4] There was a trend toward an interaction between more helpful group and evaluated group, but it did not quite reach statistical significance; $\chi^2(1) = 3.06$, $p = .08$, $b = -.12$, $SE = .07$, $t = 1.8$.

Discussion

These results replicated the group-based sampling biases from the first experiment. More importantly, we found additional evidence that people attend to the information differently depending on whether it aligns with ingroup superiority beliefs. Overall,

[4] Due to a programming error, we were not able to conduct the corresponding analysis for Experiment 1.

participants readily noticed how the groups differed when the ingroup was de facto better, and they were fairly accurate about the magnitude of the difference (see Figure 2B). However, when the outgroup did better, hence out of sync with a belief in ingroup superiority, participants downplayed the difference and provided more inaccurate estimates, on average.

Experiment 3

In this experiment we changed the sampling context, from innocuous restaurant reviews to evaluating performances in a task mimicking an airport security scan. With this change of the sampling context, participants are also asked to directly assess how good or bad the ingroup and outgroup was. In the previous studies, the group characteristic (helpfulness) might have been interpreted as secondary (while the primary information, the restaurant reviews, had nothing to do with the groups per se). We also took the experiment online and changed the cultural setting, from Scandinavia to the United States. These changes were motivated by an effort to broaden the scope of the findings from the two first experiments to investigate if they would generalize to (a) a different task that is more consequential, (b) another type of participant sample, and (c) another cultural context.

Method

Participants. 155 participants took part in the experiment on Amazon Mechanical Turk (MTurk). The invitation was open to Americans only, a maximum of 100 approved tasks in the past, and an approval rate of at least 98%. We used a restrictive setting for previous task experiences because we were concerned that non-novice MTurkers would be familiar with minimal group experiments. We recruited a much larger sample in this experiment because we observed a weaker Valence × Sampled Group interaction in Experiment 2, and because we expected the MTurk data to be noisier than our lab experiment data.

Analyses were limited to the 140 participants who answered our dependent measures (i.e., sampled at least one piece of information and estimated group performance scores), and who correctly remembered their group membership. The median age was 31 years ($SD = 10.54$), 57% were women. Participants received $1.50 for completing the experiment (*Mdn* completion time = 12 min).

Procedure and materials. This online experiment was setup in Qualtrics, with imported values for the sampling task from a database on webtask.io. The experiment started with a minimal group assignment using the same (but translated) materials and procedures as in Experiments 1 and 2. After the group manipulation, participants were introduced to a perceptual performance task in which they were asked to confirm/disconfirm the existence of weapons in x-ray images for a series bags. The similarity to an airport security screening was explicitly pointed out, and participants were asked to do the task as if other people's security depended on it. They were further instructed to work as fast as they could without compromising accuracy. The weapon screening task had 10 trials, plus a practice trial, with 20 s to respond to each.

Next, we introduced the sampling task. We informed participants that we had built a database for performances by previous participants in the weapon screening task. We further told them that we were interested in whether they could intuit how well people did on the task, based on examining the results of those who belong to either the J or P category. The rest of the instructions were the same as in Experiments 1 and 2, except for some additional information about the scores to review. Scores varied between 0 and 100, and were described as primarily capturing accuracy (% correct), but also speed and type of errors (misses would be considered worse than false alarms). This was provided to give greater salience to negative information (bad performances), as low scores would imply a danger to others in this context.

Participants could sample up to 100 scores (though they were not told of any limit), facing a choice between the J and P category each time, as in Experiments 1 and 2. Integer scores were randomly drawn from two approximate normal distributions for the respective categories. Both distributions had a standard deviation of 10, and the J distribution had a mean of 63 whereas the P distribution had a mean of 68 (i.e., Cohen's $d = .5$, as in Experiment 2). In other words, participants assigned to J belonged to a worse performing group, whereas those assigned to P belonged to a better group. Practically speaking, when participants clicked on one category, Qualtrics retrieved random numbers (with replacement) from the corresponding group dataset on webtask.io, except for the first score. We manipulated the first score to either 79 (positive first sample) or 51 (negative first sample). Once participants chose to abort the sampling, they were asked to estimate the average performance in each group. As before, the order of these group estimate questions was counterbalanced.

Finally, participants answered a series of questions about age, gender, and GHP group membership/identification (see also Experiment 1 for details). Participants were also asked to estimate how many studies in psychology and economics they had done on MTurk in the past, as well as the total number of studies. In keeping with our restrictive inclusion criteria, participants were unexperienced with studies in psychology (*Mdn* count = 5), as well as economics (*Mdn* = 2).

Results

As in the two previous experiments we first investigated which group participants began to sample from. Replicating both experiments we found a very large proportion, 92%, of participants picked the first sample from their ingroup, binomial test $H_0 = .5$: $p < .001$, 95% CI [.87, .96]. Next, we investigated if, as in the Experiments 1 and 2, participants would seek more information from their own group and possible effects of valence of the first sample and better performing group on sampling behavior. This was done by fitting a Poisson generalized linear mixed effects model with valence of first sample (positive vs. negative), better performing group (ingroup vs. outgroup) and sampled group (ingroup vs. outgroup) as fixed factors and subject as random factor. We used a Poisson model here, as opposed to a negative binomial model in the previous studies, because there was no indication of overdispersion in the data. Replicating the findings of Experiments 1 and 2, this analysis revealed a significant main effect of sampled group, $\chi^2(1) = 25.0$, $p < .001$, $b = -.15$, $SE = .03$, $z = 5.2$. Participants sampled significantly more from the ingroup, $M = 4.90$, $SD = 6.82$, than the outgroup, $M = 3.66$, $SD = 7.3$. Further, this analysis also revealed a significant Sampled Group × Valence interaction, $\chi^2(1) = 7.53$, $p = .006$, $b = -.08$, $SE = .03$, $z = 2.66$.

This interaction indicated that participants in both the positive valence condition, *M*s = 4.43 and 2.92, $t(70) = 5.10$, $p < .001$, and in the negative valence condition, *M*s = 5.13 and 4.43, $t(68) = 2.69$, $p = .009$, sampled more from the ingroup than the outgroup, but that the difference was larger in the positive valence condition.

Similar to Experiments 1 and 2, we asked participants to estimate average group scores in the sampled materials, in this case weapon screening performances. We tested possible effects of better performing group and valence of the first sample on these judgments by means of a generalized linear mixed model with valence of first sample (positive vs. negative), better performing group (ingroup vs. outgroup), and evaluated group (ingroup vs. outgroup) as fixed factors and subject as random factor.

This analysis showed three significant effects; all other *p*s > .15. First, replicating the two previous studies, there was an interaction between better performing group and evaluated group; $\chi^2(1) = 16.32$, $p < .001$, $b = 2.19$, $SE = .49$, $t = 4.5$. The interaction is illustrated in Figure 2D. As in Experiment 2, when participants belonged to the better performing group they were fairly accurate about the direction and magnitude of the difference, *M*s = 66.3 and 60.1; $t(60) = 4.24$, $p < .001$. However, when participants belonged to the worse performing group the difference between the two groups was considerably reduced, *M*s = 63.0 and 65.2; $t(78) = 1.56$, $p = .12$.

Further, there was a main effect of valence, $\chi^2(1) = 12.68$, $p < .001$, $b = 2.39$, $SE = .65$, $t = 3.66$. Participants in the positive valence condition gave higher evaluated scores, $M = 66.04$, $SD = 10.58$, than did participants in the negative valence condition, $M = 61.33$, $SD = 9.12$. The main effect of valence was, however, qualified by a Valence × Evaluated Group interaction, $\chi^2(1) = 12.51$, $p < .001$, $b = -1.66$, $SE = .49$, $t = 3.41$. The interaction is illustrated in Figure 2E. Similar to the two previous experiments, this interaction reflected that participants in the positive valence condition evaluated the ingroup as being better, $t(70) = 3.23$, $p = .002$, than the outgroup, *M*s = 68.4 and 63.7, respectively, while this difference, $t(68) = 1.19$, $p = .24$, was much smaller in the negative valence condition, *M*s = 60.5 and 62.2, respectively.

Finally, as in Experiment 2, we also submitted the signed difference between the estimated average group scores and the average individually observed scores to the same analysis as the original estimated average group scores. This analysis revealed a significant main effect of valence, $\chi^2(1) = 6.82$, $p = .009$, $b = -1.20$, $SE = .51$, $t = 2.30$; all other *p*s > .11. Participants in the negative valence condition gave an estimated average group score that slightly overestimated the observed scores, $M = 0.77$, $SD = 6.08$, while the estimated average group score in the positive valence condition slightly underestimated the observed scores, $M = -1.84$, $SD = 9.14$.

Discussion

Using a larger sample and new sampling setting in this experiment, we replicated all the findings from the previous experiments. While the sampled information in Experiments 1 and 2 might be described as trivial, we found the same effects in a context mimicking an airport security scan—a scenario where negative performances should also carry more weight. This experiment further illustrates the robustness of the effects across cultural contexts (Scandinavia vs. United States), experiment population (student vs. convenience sample), and test context (lab vs. online).

Experiment 4

In the final experiment we examined sampling related to a characteristic associated with adverse real-life stereotyping—namely, group differences in cognitive abilities (see, e.g., Devine & Elliot, 1995). As before, we examined group impression formation in a novel (minimal) group situation, and participants sampled from two group distributions, teased apart to examine the same motivational forces as in Experiments 2 and 3. The concluding experiment also comprised a preregistered replication of the principal hypotheses (see https://osf.io/c5qk4/ or http://aspredicted.org/blind.php?x=yv5fb8 for protocol details).

Method

Participants. 168 participants took part in the experiment on MTurk. We used the same inclusion criteria as in Experiment 3. Analyses included the 142 participants who answered our dependent measures (i.e., sampled at least one piece of information and estimated group performance scores), and who correctly remembered their group membership. The median age was 29 years ($SD = 10.10$), 59% were women. Participants received $2 for completing the experiment (*Mdn* completion time = 21 min). We recruited a sample of a similar size as in Experiment 3, based on that experiment having a power of at least .90 for the hypothesized effects.

Procedure and materials. The experiment was conducted online and closely mimicked Experiment 3, except that we changed the sampling context to assessing cognitive performances. In this experiment, we also introduced the sampling task right after the minimal group manipulation (unlike Experiment 3). Respondents were informed that the scores they would review reflected performances in puzzles and problem-solving tasks by other participants. They were also told that they would do the same test themselves afterward (see below). Participants were further informed that most people score between 70 and 130 (as with IQ scores), and that we were interested in where they thought a typical GHP-J and GHP-P person would fall along this continuum.

After the first sampling instance, scores were drawn at random from two normal distributions with means of 102 (Group J) and 107 (Group P) respectively. Both distributions had a standard deviation of 10 (i.e., Cohen's *d* for the actual group difference was .5, as in the previous experiments). In other words, participants assigned to J belonged to a worse performing group, whereas those assigned to P belonged to a better group. The means were slightly offset from 100 to avoid guessing exactly at the midpoint of normal range of scores, as mentioned to participants. Again, we manipulated the first sample to either be a strong performance (118, a positive first sample) or a weak one (92, a negative first sample).

After the sampling task, participants were themselves introduced to a series of puzzles and problem-solving tasks, in the form of the 16-item International Cognitive Ability Resource (ICAR) sample test (Condon & Revelle, 2014). After the test, they were asked how they think they had performed. Using a slider from 0 to

100, they indicating the percentage of all participants they thought they had performed better than. These data were collected to map individual differences in ingroup biases, but are peripheral to the current inquiry and reported separately (article in preparation). All other aspects of the experiment were identical to Experiment 3.

Results

We expected participants to begin sampling from their ingroup. As in the three previous experiments, we again found a very large proportion of participants, 92%, picked the first sample from their ingroup, binomial test $H_0 = .5$: $p < .001$, 95% CI [.87, .96].

By fitting a Poisson generalized mixed effects model, with valence of first sample (positive vs. negative) better performing group (ingroup vs. outgroup) and sampled group (ingroup vs. outgroup) as fixed factors and subject as random factor, we next investigated if participants sought more information from their own group, and possible effects of valence of the first sample and better performing group on their sampling behavior. This analysis revealed three significant effects, all other *p*s $> .21$.

First, there was a significant main effect of Sampled group, replicating Experiments 2 and 3, $\chi^2(1) = 82.4$, $p < .001$, $b = -.28$, $SE = .03$, $z = 8.7$. Participants sampled more from the ingroup, $M = 5.07$, $SD = 6.17$, than the outgroup, $M = 2.93$, $SD = 4.61$. Second, there was a significant Sampled Group × Better Performing Group interaction, $\chi^2(1) = 6.86$, $p = .008$, $b = .09$, $SE = .031$, $z = 2.86$. Participants in the better ingroup condition sampled slightly more from the ingroup, $M = 5.87$, than participants in the condition with the outgroup performed better, $M = 4.29$, while the number of samples taken from the outgroup was similar in both conditions, *M*s = 2.94 and 2.91. Finally, the analysis also revealed a significant main effect of valence of the first sample, $\chi^2(1) = 4.4$, $p = .036$, $b = .18$, $SE = .08$, $z = 2.3$. Participants in the positive valence condition sampled more in total, $M = 4.79$, $SD = 6.7$, than participants in the negative valence condition, $M = 3.17$, $SD = 3.87$.

After having terminated the sampling, participants in Experiment 4 judged the performance of members of the in- and outgroup on the puzzles and problem-solving task. We tested possible effects of better performing group and valence of the first sample on these judgments by fitting a generalized linear mixed model with valence of first sample (positive vs. negative), better performing group (ingroup vs. outgroup), and evaluated group (ingroup vs. outgroup) as fixed factors and subject as random factor.

This analysis revealed two significant effects; all other *p*s $> .11$. We found a significant Better Performing Group × Evaluated Group interaction; $\chi^2(1) = 6.14$, $p = .01$, $b = 1.58$, $SE = .63$, $t = 2.5$. The interaction, illustrated in Figure 2F, replicated the pattern of results from Experiments 2 and 3 to show that the rated difference in performance between the two groups was fairly accurate when participants themselves belonged to the better group, *M*s = 103.7 and 98.4, $t(68) = 2.76$, $p = .007$. In contrast, there was a downplayed difference in judgments of the two groups when participants belonged to the worse performing group, *M*s = 101.8 and 102.8, $t(71) = 0.62$, $p = .54$.

As in Experiment 3, this analysis also revealed a main effect of valence, $\chi^2(1) = 8.68$, $p = .003$, $b = 2.07$, $SE = .69$, $t = 3.02$. Participants in the positive valence condition gave an overall rated performance on the cognitive test that was higher, $M = 103.7$, $SD = 10.11$, than did participants in the negative valence condition, $M = 99.6$, $SD = 12.4$. All other *p*s $> .11$ except for the Better Performing Group × Valence interaction, $\chi^2(1) = 3.35$, $p = .067$, $b = 1.24$, $SE = .69$, $t = 1.81$. The Valence × Evaluated Group interaction was not replicated here in terms of significance ($p = .21$), but the trend was the same as in the other studies—that the ingroup rating among those receiving a positive first sample was slightly higher than all other ratings (see Figure 2G).

Next we examined how participants' evaluations might deviate from the data they had experienced in their samples. Accordingly, we submitted the signed difference between the estimated average group scores and the average individually observed scores to the same analysis as the original estimated average group scores. This analysis revealed no significant effects, all *p*s $> .12$.

Discussion

This experiment replicated all the previously documented effects, with one exception, using problem-solving abilities (intelligence) in the sampling task. The interaction between first sample and evaluated group (ingroup vs. outgroup) was not significant here, but the trend was in line with that in all the other studies—that those receiving initially positive information about the ingroup overestimated the ingroup relative to the other group. A potential reason for a weaker effect here is that first impressions weight less heavily when people have more a priori knowledge about the characteristic in itself. Presumably, most people have less knowledge about the ability to detect weapons in x-ray images, for themselves and others, as compared to their knowledge about cognitive abilities. This opens up for more self-anchoring (projecting knowledge about oneself onto beliefs about one's own group; Cadinu & Rothbart, 1996), which in turn could make ingroup evaluations less sensitive to the impact of novel information about other ingroup members. Another difference between this experiment and the previous one is that participants in this sampled from both the ingroup and outgroup when the first sample was positive. More broadly, there were no robust (replicated) effects of valence on sampling behaviors across studies.

Participants gravitated somewhat toward a score of 100 in their group evaluations, overall, which would be consistent with a priori knowledge about average IQ scores. It is also consistent with general anchoring effects in numbers (Rosch, 1975), and guessing in the middle of the mentioned range of scores. All these things considered though, most patterns in the group evaluations (as well as the sampling behaviors) were highly consistent with the findings from the previous studies. For example, just like in Experiments 2 and 3, participants were quite perceptive about the actual group difference when their own group fared better, while those in the worse performing group downplayed the difference.

The Role of Experience—Auxiliary Analyses

In each experiment, we tested hypotheses about sampling behavior and group evaluations, respectively. We did not, however, explore how the two outcomes are related, and we did not model that participants would de facto have different experiences of the groups, both by chance and as a function of the manipulations. Here we provide such tests, positing individual experiences with the ingroup and outgroup as mediators of the experimental effects and sampling behaviors, and their interactions, on the group evaluations. The reason for presenting these tests here, rather than for each experiment, is primarily that these individual difference analyses would be underpowered in each separate experiment. We therefore combine the results from Experiments 3 and 4 in a multigroup (group = experiment) path analysis. We focus on the last two experiments because (a) Experiment 1 had a different design, (b) the first two experiments were too small for testing this kind of model, and (c) because all the design features and metrics were the same in Experiments 3 and 4 (unlike Experiment 2). Adding a constant of 39 to the group means in Experiment 3 makes them equivalent to those in Experiment 4, and hence easy to analyze together after group (experiment) centering.

Rationales

A few concrete rationales guided these analyses. First, as described in the introduction, first sample valence should impact ingroup experiences only, among those who chose to sample from the ingroup first. Thus, there should be an interaction between the manipulation of the valence of the first sample and the first chosen group in predicting overall ingroup experiences. Because so many participants sampled from the ingroup first, this interaction should also drive a main effect of the first sample manipulation on ingroup experiences. Second, the first sample effect should get weaker the more information people gather (i.e., with more information, those with initial positive or negative information will regress toward the true mean of the ingroup). Third, as an omnibus effect, ingroup experiences should be more systematic overall, as compared to outgroup experiences, due to motivated (biased) sampling. An unbiased sampler would be equally likely to start sampling from the outgroup, and hence over- or underestimate that group instead. Concretely, the explained variance of all manipulations, sampling behaviors, and their interactions, should be higher in the ingroup experiences if, and only if, there was asymmetric (motivated) sampling from the ingroup relative to the outgroup. Fourth, in line with a data-driven process and (nonmotivational) sampling models, we acknowledge that individual experiences should be a proximate predictor of a person's evaluations of the two groups. As such, we also expected ingroup experiences to mediate the evaluative effects of the manipulations and sampling behaviors. Finally, we allowed for the possibility that people anchor outgroup evaluations based on ingroup experiences, and vice versa, but we had no predictions about the strength of these cross-over effects.

Model Specification

Because we focused here on interactions between the manipulations and the sampling behaviors, and ingroup and outgroup experience as mediators, we positioned the sampling behaviors as independent variables. A more realistic model, with number of samples as (moderating) mediators, cannot be reliably estimated due to their distribution (i.e., there are no formal models for testing mediation with count data variables; see, e.g., Muthén, 2014). Thus, the model had 15 independent variables, reflecting the manipulations (better/worse group, valence of first information), sampling decisions (first choice, number of ingroup and outgroup samples), and two- and three-way interactions between these. The full model is illustrated in Figure S1 in the materials available for download at https://osf.io/c5qk4/.

Limiting the complexity of the model, we did not include the four-way interactions between the product of the manipulations, first choice, and ingroup or outgroup samples. We also imposed the natural limitation that the amount of ingroup sampling could only influence ingroup experiences, and outgroup sampling affecting outgroup experiences. The dependent variables were ingroup and outgroup ratings, and personal overall experiences with the ingroup and outgroup were mediators. Mediation effects were estimated with bootstrapping (5,000 draws), and the model was estimated in Mplus 7.3 (Muthén & Muthén, 2012).

Model Estimation and Results

In our multigroup framework, we first estimated a model in which we assumed all effects to be equal across Experiments 3 and 4. This way we could also formally test for variations across studies, and estimate average effects (similar to a mini meta-analysis). We planned to free the different experiment estimates for particular paths based on indications of model misfit, but results indicated excellent fit of the constrained model; $\chi^2(86) = 71.29$, $p = .87$, comparative fit index = 1.00, root mean square error of approximation = .00, 90% CI [.00, .03], standardized root mean square residual = .05. For illustrative purposes, we show main effects and relevant first-order interactions in Figure 3. The higher order interactions were not significant, and the lower level better/worse interactions were only relevant for estimating those. Complete results are presented in the materials available for download at https://osf.io/c5qk4/.

It is not surprising how the manipulations impact participants' experiences. For instance, those in the better condition should have more positive ingroup experiences, and less positive outgroup experiences. It is also sensible that those experiences in turn predict how people evaluate the groups: Those who had more positive experiences with the ingroup or outgroup, as a function of the first sample, also evaluated that group more positively, and the same goes for the outgroup evaluations. In other words, people provide estimates of group attributes in line with what they have experienced, but failed to adjust for the errors inherent in their small samples (see also Fiedler, 2000). What is more important for our purposes, however, is that we also have clear evidence that the biased experiences depend on the chosen (motivated) behaviors. The valence manipulation indeed affects ingroup experiences only, while there should have been an equivalent effect for outgroup experiences if people made an unbiased (random) choice of where to start sampling. Underpinning that conclusion, we also found the critical interaction effect between the valence manipulation and first choice for the ingroup experiences. The indirect effect on ingroup evaluations was also significant ($B = 5.13$, $p = .004$, 95% CI [1.64, 8.62]). In contrast, the better/worse group manipulation had largely symmetric, inverse effects on ingroup and outgroup experiences: Those who belonged to the better group naturally had better ingroup experiences and worse outgroup experiences. This is what to expect from a manipulation not interacting with the sampling choices overall.

We further predicted that the first sample manipulation effect would get weaker as people sample more, but we could not empirically substantiate that notion here. Potentially, this is due to

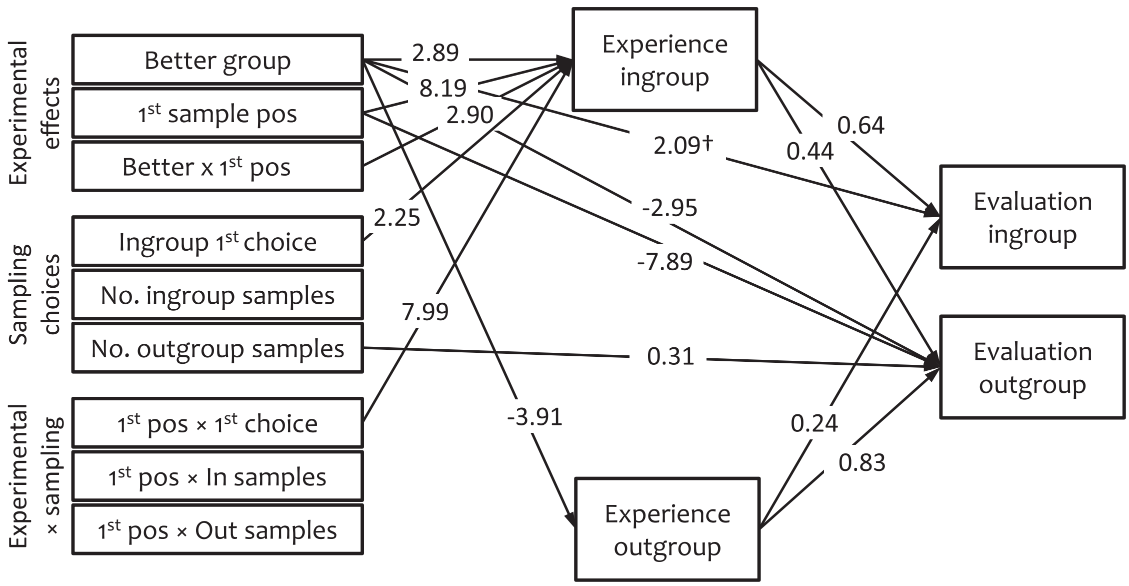

Figure 3. Path diagram illustrating significant ($p < .05$, except † = .051) unstandardized effects of the manipulation, sampling choices, and theory-relevant interactions on group evaluations, as partially mediated by overall experiences with the ingroup and outgroup. Coefficients represent average effects in Experiments 3 and 4. Pos = positive; No. = number.

most participants sampling relatively little, and to undue the effect of the first sample a lot more sampling would be needed, among many more participants. Theoretically, the valence effect should dominate with small amounts of information, whereas the better/worse effect should be become more pronounced with more sampling. In line with this notion, ingroup experiences mediated the effect of the valence manipulation ($B = 5.26$, $p < .001$, 95% CI [2.8, 7.63]). The better/worse group manipulation did so as well, but to a lesser extent ($B = 1.86$, $p = .002$, 95% CI [0.67, 3.05]). The indirect effect for the interaction of the two manipulations was marginally significant ($B = 1.86$, $p = .07$, 95% CI [−0.16, 3.88]).

At the omnibus level, we also found evidence that sampling choices influence overall experiences with the groups. The independent variables together explained 63% of the variance in participants' experiences with their ingroup, but only 12% of the outgroup experiences. Although the manipulations by themselves contribute to the predictability of ingroup experiences (and equally so for the outgroup experiences in the case of the better/worse manipulation), only asymmetric sampling behaviors from ingroups and outgroups could reasonably account for that difference. If people were sampling in an unbiased fashion the explained variance for ingroup and outgroup experiences would be equivalent. They were clearly not.

Finally, the model also included direct effects for the outgroup evaluations, which we had not specifically hypothesized. For example, those who started sampling from the ingroup had somewhat more positive experiences of that group. That could be a chance finding, or it could indicate that those who started with the ingroup also were more inclined to terminate the sampling process with positive ingroup information. As experiences were random, net of the manipulations (and those represented symmetric deviations from the grand mean), there should not have been an effect otherwise. Further, those who belonged to the better group rated the outgroup lower, even after accounting for different experiences, and so did those who received initial positive information. Those in the better group condition also rated the ingroup higher, net of experiential effects. Taken together, this suggests that those who receive positive information about their ingroups, in different forms, exaggerated the difference to the outgroup beyond what the data suggest, especially in response to the better/worse manipulation. Finally, those who sampled more from the outgroup also evaluated it more positively.

General Discussion

Studying sampling behaviors in a minimal group situation, this article illustrates a tandem operation of preexisting motivations and experience in shaping impressions of novel groups. It is intuitive that both motivation and experience should play a role, but research has primarily focused on either explanation in its own right (see, e.g., Denrell, 2005; Tajfel et al., 1971). Some studies have examined how motivation moderate inferences about groups, but these have employed fixed sets of stimuli (e.g., Howard & Rothbart, 1980; Hughes et al., 2017; Schaller & Maass, 1989), and have hence not considered that people also make choices about the information they gather. Here, we examined if the sampling behaviors themselves are biased in favor of the ingroup, and if so, what the effects of initial experiences are on impression formations when people are facing novel group categorizations.

A clear finding in all studies is that virtually everyone sought information about the ingroup first, given a choice to do so. Another robust finding was that participants chose to sample more information in total from the ingroup, as compared to the outgroup. This illustrates how the process of information seeking may itself be guided by group motives. On the other hand, when people actively gathered information about the groups there was no main effect of evaluated group (ingroup vs. outgroup), as in numerous minimal group studies (see Mullen et al., 1992). This suggests that the task of actively gathering information about groups may reduce biases overall, and that certain information may nullify evaluative biases altogether—at least in the context of novel groups.

Those who initially received negative information showed, on average, no ingroup favoritism in the group evaluations (see left-hand bars in Figure 2A, C, E, and G). Those who received initially positive information, in contrast, overestimated how good the ingroup was, in each experiment. As the auxiliary analyses further show, those who overall received more positive ingroup information naturally evaluated the ingroup more positively (and vice versa for negative ingroup experiences). In line with existing sampling models, this suggests that people can make reasonable inferences from what they have experienced, while being naïve about how their experiences are skewed in relation to the true group attributes (see also Fiedler, 2000; Fiedler & Juslin, 2006; Juslin et al., 2007). These experiences are bound to be skewed because people are not gathering enough information (i.e., if participants gathered information indefinite, their experiences would regress toward the true group means). More important for our purposes, we showed that group experiences were not randomly skewed. Instead, they were systematically skewed for the ingroup only, because of the choice to start gathering information from this group (see Figure 3). Further, the biased experiences that people came to have, due to their choices, were interpreted differently depending on whether they painted a positive or negative portrait of the ingroup. On average, participants ignored the initial negative data about the ingroup when they evaluated the groups. This interpretative bias, however, had the net result that the two groups were rated equal, and perhaps ironically, made these participants less biased by nonrepresentative data. In other words, those who received initially positive ingroup information were more rational in the sense of estimating group attributes based on what they had actually seen, but they were ultimately biased by virtue of not discounting for their nonrepresentative experiences. In contrast, those who received negative information were biased in the sense of not following the data they had seen, but less biased in the sense of (coincidentally) discounting for nonrepresentative observations. Overall, this illustrates different kinds of biases in novel group impressions, as well as different kinds of accuracies, beyond existing research on motivated cognition (e.g., Brewer, 1979; Tajfel & Turner, 1979) and experience sampling (e.g., Denrell, 2005).

Whereas previous sampling models describe how ingroup biases emerge from decisions about when to stop sampling, and erroneous inferences about outgroups (Denrell, 2005), we show how biases emerge from decision about how to start sampling, and erroneous inferences about ingroups. A critical difference compared to Denrell's model is also that it requires extensive and "accidental" information about groups, while we focused on situations where people gather little information, by choice. In reality, we would expect group evaluations to depend both on systematic initial choices and on later accidental encounters, and the effects to vary across different group contexts, as well as the overall amount of experiences with different groups. We did not find that the amount of sampling from the ingroup or outgroup influenced the group ratings, or moderating the other effects. However, with more participants and more sampling in general (implying greater statistical power), we would expect to see such effects, especially in relation to the first sample manipulations. Manipulating amounts of information to sample, while leaving open the proportions from each group, could help address this question in future work.

For our second manipulation, concerning actual differences between groups, we focused mostly on inferential biases. Participants were perceptive about actual group differences, when (and only when) the data indicated that the ingroup was better (see left-hand panels in Figure 2). Again, this suggests that people can be accurate about average characteristics of novel groups, just as they are about nonsocial categories (e.g., Lindskog & Winman, 2014; Malmi & Samson, 1983), while they selectively followed the data when estimating group attributes. Unlike the case of positive or negative initial information, the disregard of unfavorable information about actual group differences consistently led to inaccurate evaluations. Given that ingroup and outgroup experiences were similarly (albeit inversely) affected by this manipulation, this further points toward interpretive biases, rather than a data-driven bias. Taken together then, we have suggestive evidence of two kinds of biases in the information seeking process—one that has to do with failing to discount for nonrepresentative data, and another having to do with direct misinterpretations/distortion of the observed data. The first bias fits with those emphasized in the sampling literature (e.g., Fiedler, 2000), whereas the second dovetails findings in the literature on motivated group biases (e.g., Hastorf & Cantril, 1954).

Previous research has suggested that differences in evaluating self and others' performance can arise because of an asymmetry in the amount of information one has about own and others' performance (Moore & Healy, 2008). Our participants, however, sampled similarly from the two groups regardless of which group was better, and the tendency to downplay the difference when the outgroup was better cannot be attributed to differences in the amount of information. Instead, participants were seemingly less willing to see and/or report a difference when the data paints a negative portrait of the ingroup.

Implications of Group-Motivated Sampling

Some of the motivations behind these findings might have been quite subtle and innocent, such as being more curious about ingroups than outgroups. Still, the consequences are not innocuous. The consequence of people having realistic evaluations of outgroups, on average, but optimistic evaluations of ingroups when initial experiences are positive, is still a discriminatory outcome. For example, imagine a professor starting a job at a new institute, and the first master student s/he works with is excellent. What these findings suggest is that s/he may overestimate the talents of other students from that institute, relative to those from other unfamiliar institutes, when hiring graduate students. This is plausible regardless if the talent concerns a cognitive (Experiment 4), perceptual (Experiment 3), or a social skill (Experiment 1 and 2). The effect of first impressions was weaker (and nonsignificant) for intelligence in Experiment 4, but the pattern was still the same in all studies (see right-side panels in Figure 2).

While the interactions with first impressions illustrate an additional source of bias, on top of baseline ingroup favoritism, it also points toward novel ways to design prejudice interventions. Virtually all prejudice interventions to date are targeting negative evaluations of outgroups (Paluck & Green, 2009), and the most common strategy—outgroup contact—is based on simply accumulating more positive outgroup experiences (Pettigrew & Tropp, 2006). From a sampling perspective, the presumed problem is a lack of positive experiences with outgroups, so it is makes sense that the proposed solution is to seek more of those experiences (see also Denrell, 2005). However, our findings indicate that the root of the problem might have just as much to do with assigning too much weight to initially positive ingroup experiences. We know from the auxiliary experiential analyses that this bias is tied to a decision to start sampling from the ingroup, and intuitively then, one could imagine that biases would be reduced if people instead started sampling from outgroups. Unfortunately, they were too few participants who started sampling from the outgroup to directly test that notion in these studies (and we could not reliably estimate the experiential model for these observations alone). Self-selecting outgroup information could also be different from being explicitly encouraged to sample that way, and it is possible that most people would assign more weight to initial negative experiences if they started sampling from the outgroup. On the other hand, research suggests that biases are more likely to arise when people can favor ingroups in terms of positive outcomes, as opposed to putting down outgroups in terms of negative outcomes (e.g., Brewer, 1999; Mummendey & Otten, 1998). From this perspective, negative initial outgroup information might be less harmful than positive initial ingroup information. In any case, the potential of reducing prejudice by focusing on how people gather and process ingroup information, and especially early experiences, is worth exploring as a complement to the many interventions that all focus on outgroups and later experiences.

It is also worth noting that the valence of initial information is considered highly relevant in individual impression formation (e.g., Peeters & Czapinski, 1990; Skowronski & Carlston, 1989), and what we have done here is to extend the argument to learning processes about groups overall, with minimal information about individual group members. Unlike extensive research on individual impression formation, however, we did not find a negativity bias, indicating negative information carries more weight that positive information. The current findings indicating greater weight for positive ingroup information is especially noteworthy when considering arguments such as the following: "Hardly any exceptions (indicating greater power of good) can be found [. . .] bad is stronger than good, as a general principle across a broad range of psychological phenomena" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 323). Still, more recent analyses suggest that learning about group members seems to present a general class of exceptions and caveats to that principle (see Hughes et al., 2017; Paolini & McIntyre, 2019). For clarification, our argument here is not that positive information always outweighs negative information in group situations. Still, it is reasonable that negative information carry more weight in interpersonal interactions, as compared to when people learn about groups as a whole, especially when the group members are anonymous—as in our experiments. Reasons for paying more attention to negative information, such as assessing trust or threats, are presumably less salient without direct interactions. That does not mean that our kind of group situation is nonexistent. Especially in online forums, there are ample opportunities to sample information about groups from anonymous members. Our experiments suggest that people's attention to positive and negative information might not be quite as universal as previously proposed, but varying with the features of the social situation.

Caveats and Future Directions

The exploration of sampling behavior from novel experimental groups is informative to describe basic principles for social information seeking, especially in the absence of preexisting stereotypes and extensive experiences with various groups. To some scholars, the absence thereof may seem like a limitation for the generalizability of these findings. Without negating that the real-life implications of these findings call for more attention, we do believe there are plenty of situations in which adults learn about novel groups. However, the type of groups here are likely to be different from the types of groups that prejudice researchers typically study. Prejudice research mostly examines groups that are best described as broad social categories, encompassing thousands, or millions of people, such as Black and Hispanic Americans, Jews, Gays, and so on (e.g., Dovidio & Gaertner, 2010; Sibley & Barlow, 2016). However, when people move, change jobs, start educations, join teams or clubs, etcetera, they do encounter novel social categorizations. These types of groups are not only common (e.g., Forsyth, 2018), but typically also considered among the most entitative and important in people's lives (Lickel et al., 2000). Furthermore, if we take serious that experience associated with a common ingroup can reduce prejudice (e.g., Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993), we should also take serious how experiences with new people may catalyst prejudice when large ingroups split into subgroups. These remarks aside, however, it remains a valid point that the learning we have examined might differ substantially from how people learn about new characteristics, or new exemplars, from a previously known social category.

When participants were evaluating new groups on a more familiar trait (problem-solving abilities), they seemed to attach less weight to the first sample, but showed the same interactions as before with belonging to a better or worse performing group. Potentially, that interaction also extends to novel information gathering situations, involving unfamiliar members, but familiar groups and a familiar characteristic. For example, consider a White person in a hiring committee, reviewing a sample of White and Black applicants. If the current sampling behaviors extend to gathering information about well-known groups, chances are that s/he will be more inclined to start with a White application and seek out more information about this pool overall. These findings further highlight the possibility that a single positive instance in the beginning may bias him/her to overestimate White applications overall. Examining if that is the case, and studying sampling behaviors more broadly with well-known groups, and in this presence of stereotypes, would be a natural and important extension of this work.

With novel groups, there are also other factors besides group membership that are likely to influence sampling behaviors, and led to evaluative biases. For example, something that is likely to be present in the navigation of many (if not most) novel group situations in real-life, are status differences. Members of high status groups spontaneously show greater ingroup biases (e.g., Nosek, Banaji, & Greenwald, 2002), also in minimal groups (Mullen et al., 1992). A natural question then is whether members of high and low status groups equally likely to seek out more information about their own group. Humans and other primates often show particular interest in powerful and high status individuals (e.g., Chance, 1967; Fiske, 1993a; Shepherd, Deaner, & Platt, 2006), and more information sampling related to these could potentially underpin status-asymmetric ingroup–outgroup evaluations. More broadly, sampling behaviors could potentially provide a unifying framework for studying different kinds of group biases. For instance, social psychologists have traditionally suggested that ingroup biases and status-based biases are based on separate motivations (e.g., Jost & Banaji, 1994; Sidanius & Pratto, 1999). An alternative (or perhaps complementary) explanation is that these biases manifest themselves as independent because people sample information along two separate dimensions, while a common motivation underpin both, in terms of wanting to learn more about certain groups (ingroups and high status groups), relative to other groups (outgroups and low status groups). Or perhaps sampling behaviors are different along the status dimension from what we have observed about novel ingroups and outgroups. In any case, examining the similarities and differences in sampling behaviors for different kinds of group biases would seem a relevant topic for future research.

Concluding Remarks

Category learning and impression formation are classic topics in psychological research (e.g., Ashby & Maddox, 2005; Brewer, 1988), and yet we know little about how early experiences with individual group members influence the overall representations of ingroups and outgroups. These studies illustrate how beliefs about ingroup positivity interact in meaningful ways with experiences that confirms or disconfirms those beliefs, with results than range from fairly accurate to clearly biased beliefs about ingroups and outgroups.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481. http://dx.doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. http://dx.doi.org/10.1146/annurev.psych.56.091103.070217

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 883–885. http://dx.doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Bergh, R., Akrami, N., Sidanius, J., & Sibley, C. G. (2016). Is group membership necessary for understanding generalized prejudice? A re-evaluation of why prejudices are interrelated. *Journal of Personality and Social Psychology, 111,* 367–395. http://dx.doi.org/10.1037/pspi0000064

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

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Skrull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

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

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology, 70,* 661–677. http://dx.doi.org/10.1037/0022-3514.70.4.661

Chance, M. R. A. (1967). Attention structure as the basis of primate rank orders. *Man, 2,* 503. http://dx.doi.org/10.2307/2799336

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence, 43,* 52–64. http://dx.doi.org/10.1016/j.intell.2014.01.004

Dawtry, R. J., Sutton, R. M., & Sibley, C. G. (2015). Why wealthier people think people are wealthier, and why it matters: From social sampling to attitudes to redistribution. *Psychological Science, 26,* 1389–1400. http://dx.doi.org/10.1177/0956797615586560

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112,* 951–978. http://dx.doi.org/10.1037/0033-295X.112.4.951

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton Trilogy revisited. *Personality and Social Psychology Bulletin, 21,* 1139–1150. http://dx.doi.org/10.1177/01461672952111002

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1084–1121). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470561119.socpsy002029

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences, 12,* 248–253. http://dx.doi.org/10.1016/j.tics.2008.04.006

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development, 82,* 793–811. http://dx.doi.org/10.1111/j.1467-8624.2011.01577.x

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87,* 293–311. http://dx.doi.org/10.1037/0022-3514.87.3.293

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103,* 193–214. http://dx.doi.org/10.1037/0033-295X.103.1.193

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107,* 659–676.

Fiedler, K., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–32). Cambridge, United Kingdom: University Press.

Fiske, S. T. (1993a). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48,* 621–628. http://dx.doi.org/10.1037/0003-066X.48.6.621

Fiske, S. T. (1993b). Social cognition and social perception. *Annual Review of Psychology, 44,* 155–194. http://dx.doi.org/10.1146/annurev.ps.44.020193.001103

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. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 1–74). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/S0065-2601(08)60317-2

Forsyth, D. R. (2018). *Group dynamics* (7th ed.). Boston, MA: Cengage Learning.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology, 4,* 1–26. http://dx.doi.org/10.1080/14792779343000004

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases in judgments of social environments. *Psychological Science, 23,* 1515–1523. http://dx.doi.org/10.1177/0956797612445313

Galesic, M., Olsson, H., & Rieskamp, J. (2013). False consensus about consensus. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 472–476). Austin, TX: Cognitive Science Society.

Gramzow, R. H., Sedikides, C., & Gaertner, L. (2001). Memory for in-group and out-group information in a minimal group context: The self as an informational base. *Journal of Personality and Social Psychology, 80,* 188–205. http://dx.doi.org/10.1037/0022-3514.80.2.188

Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49,* 129–134. http://dx.doi.org/10.1037/h0057880

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13,* 517–523.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology, 38,* 369–425. http://dx.doi.org/10.1146/annurev.ps.38.020187.002101

Howard, J. W., & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of Personality and Social Psychology, 38,* 301–310. http://dx.doi.org/10.1037/0022-3514.38.2.301

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience, 12,* 49–60. http://dx.doi.org/10.1093/scan/nsw147

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,* 1–27. http://dx.doi.org/10.1111/j.2044-8309.1994.tb01008.x

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106,* 259–298. http://dx.doi.org/10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132,* 133–156. http://dx.doi.org/10.1037/0096-3445.132.1.133

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114,* 678–703. http://dx.doi.org/10.1037/0033-295X.114.3.678

Kalish, C. W., Rogers, T. T., Lang, J., & Zhu, X. (2011). Can semi-supervised learning explain incorrect beliefs about categories? *Cognition, 120,* 106–118. http://dx.doi.org/10.1016/j.cognition.2011.03.002

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 131,* 287–297. http://dx.doi.org/10.1037/0096-3445.131.2.287

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228. http://dx.doi.org/10.1037/0033-295X.94.2.211

Krause, F., & Lindemann, O. (2014). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods, 46,* 416–428. http://dx.doi.org/10.3758/s13428-013-0390-6

Kteily, N., Cotterill, S., Sidanius, J., Sheehy-Skeffington, J., & Bergh, R. (2014). "Not one of us": Predictors and consequences of denying ingroup characteristics to ambiguous targets. *Personality and Social Psychology Bulletin, 40,* 1231–1247. http://dx.doi.org/10.1177/0146167214539708

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology, 68,* 565–579. http://dx.doi.org/10.1037/0022-3514.68.4.565

Kutzner, F., & Fiedler, K. (2016). Stereotypes as pseudocontingencies. *European Review of Social Psychology, 28,* 1–49. http://dx.doi.org/10.1080/10463283.2016.1260238

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39,* 329–358. http://dx.doi.org/10.1207/s15327906mbr3902_8

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review, 118,* 379–392. http://dx.doi.org/10.1037/a0023010

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology, 78,* 223–246. http://dx.doi.org/10.1037/0022-3514.78.2.223

Lindskog, M., & Winman, A. (2014). Are all data created equal? Exploring some boundary conditions for a lazy intuitive statistician. *PLoS ONE, 9,* e97686. http://dx.doi.org/10.1371/journal.pone.0097686

Lindskog, M., Winman, A., & Juslin, P. (2013). Naïve point estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 782–800. http://dx.doi.org/10.1037/a0029670

Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning and Verbal Behavior, 22,* 547–559. http://dx.doi.org/10.1016/S0022-5371(83)90337-7

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502–517. http://dx.doi.org/10.1037/0033-295X.115.2.502

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology, 22,* 103–122. http://dx.doi.org/10.1002/ejsp.2420220202

Mummendey, A., & Otten, S. (1998). Positive–negative asymmetry in social discrimination. *European Review of Social Psychology, 9,* 107–143. http://dx.doi.org/10.1080/14792779843000063

Muthén, B. (2014, May). *Re: Testing a mediation with count data* [Online forum comment]. Retrieved from http://www.statmodel.com/discussion/messages/23/6872.html?1514068843

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6,* 101–115. http://dx.doi.org/10.1037/1089-2699.6.1.101

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631. http://dx.doi.org/10.1037/0033-295X.101.4.608

Pachur, T., Hertwig, R., & Rieskamp, J. (2013). Intuitive judgments of social statistics: How exhaustive does sampling need to be? *Journal of Experimental Social Psychology, 49,* 1059–1077. http://dx.doi.org/10.1016/j.jesp.2013.07.004

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology, 60,* 339–367. http://dx.doi.org/10.1146/annurev.psych.60.110707.163607

Paolini, S., & McIntyre, K. (2019). Bad is stronger than good for stigmatized, but not admired outgroups: Meta-analytical tests of intergroup valence asymmetry in individual-to-group generalization experiments. *Personality and Social Psychology Review, 23,* 3–47. http://dx.doi.org/10.1177/1088868317753504

Peeters, G., & Czapinski, J. (1990). Positive–negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1,* 33–60. http://dx.doi.org/10.1080/14792779108401856

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90,* 751–783. http://dx.doi.org/10.1037/0022-3514.90.5.751

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–547. http://dx.doi.org/10.1016/0010-0285(75)90021-3

Schaller, M., & Maass, A. (1989). Illusory correlation and social categorization: Toward an integration of motivational and cognitive factors in stereotype formation. *Journal of Personality and Social Psychology, 56,* 709–721. http://dx.doi.org/10.1037/0022-3514.56.5.709

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Current Biology, 16,* R119–R120. http://dx.doi.org/10.1016/j.cub.2006.02.013

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. New York, NY: Houghton Mifflin.

Sibley, C. G., & Barlow, F. K. (2016). *Cambridge handbook on the psychology of prejudice*. New York, NY: Cambridge University Press.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139175043

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105,* 131–142. http://dx.doi.org/10.1037/0033-2909.105.1.131

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212. http://dx.doi.org/10.1037/0022-3514.36.11.1202

Stelzl, M., Janes, L., & Seligman, C. (2008). Champ or chump: Strategic utilization of dual social identities of others. *European Journal of Social Psychology, 38,* 128–138. http://dx.doi.org/10.1002/ejsp.446

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1,* 149–178. http://dx.doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Pacific Grove, CA: Brooks/Cole.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110. http://dx.doi.org/10.1037/h0031322

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology, 12,* 129–140. http://dx.doi.org/10.1080/17470216008416717

Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: In-group love or out-group hate? *Evolution and Human Behavior, 30,* 229–237. http://dx.doi.org/10.1016/j.evolhumbehav.2009.02.004

Received August 4, 2017
Revision received December 25, 2018
Accepted February 17, 2019 ■

Correction to Grossmann, Oakes, and Santos (2019)

In the article "Wise Reasoning Benefits From Emodiversity, Irrespective of Emotional Intensity" by Igor Grossmann, Harrison Oakes, and Henri C. Santos (*Journal of Experimental Psychology: General*. Advance online publication. January 28, 2019. http://dx.doi.org/10.1037/xge0000543), all references to Study 6 in the text and Table 5 should be omitted and result in the following corrections: The second sentence in the Research Overview section should appear instead as "To compare the effects of emodiversity and emotional intensity, in Study 2, we simultaneously entered them as predictors of wise reasoning." The seventh sentence in the Research Overview section should be deleted. The last phrase of the Research Overview section should appear instead as ". . . and human-coded content analyses of wise reasoning in participants' narratives (Study 5)." The last sentence of the final paragraph in the Study 5: Emodiversity and Wise Reasoning About Intergroup Conflict section should appear instead as "We addressed this question with the final study". The title of Table 5 should appear instead as *Effects of Emodiversity and Emotional Intensity on Wise Reasoning in Study 5*. Some additional corrections to the text follow: The first sentence of the second paragraph of the Results and Discussion section of Study 1 should appear instead as "A chi-square test on the counts of affective themes in the text corpora indicated a significant difference between groups". The fifth sentence in the Participants section of Studies 4a to 4c should appear instead as "Demographics and final sample sizes are presented in Table 2." The last sentence in the Participants section of Studies 4a to 4c should appear instead as "All exclusions are reported in Table 2." The last sentence of the Participants section of Study 5 should appear instead as "We report further demographics in Table 2". The right panel of Figure 1 is missing, and the correct Figure 1 now appears in the article.

All versions of this article have been corrected.

http://dx.doi.org/10.1037/xge0000626