Changing the Gender-Science Stereotype: The Effects of Pseudocontingencies on Attenuating or Upholding Social Stereotypes.

Rianne Hop (5657903)

Utrecht University

First supervisor: Chris Harris

Second supervisor: Hans Marien

Date: 23-06-2020

Word-count: 7884

Publicly accessible: yes

Abstract

The aim of the current study was to test whether initial biases about social groups would be maintained during an extensive sampling period. Specifically, the study examined whether initial biases regarding the gender-science stereotype, that is, the bias that science is associated with men over women, were upheld or attenuated in different environments. The current research assumes stereotypes to occur due to associations between attributes and social groups and it was expected that an initial pseudocontingency-inference (PC) between gender and success or failure in solving math problems would be upheld in an environment where participants predominantly experienced success, while the same PC-inferences would be attenuated in an environment where participants experienced mostly failure. Results show that an initial bias, regarding gender and math ability, can be induced. However, contrary to the expectations, during and after the sampling phase, the initial bias was attenuated in both environments. Moreover, no difference between the pre- and the post-measure was found. Alternative explanations explore the process of subtyping and the complexity of the social world. Recommendations for future research emphasize the importance of unknown mediators or moderators and an improvement of the current study design. Practical implications regard the notion to show variations to the initial gender-science stereotype, such that women are associated with the ability to perform sciences too, in order to overcome this damaging stereotype.

Introduction

In 2018, the Nobel Prize in Physics was awarded to Professor D. Strickland, who is only the third women to win a Noble Prize since its introduction in 1901 (Lunnemann, Jensen, & Jauffred, 2019). The overrepresentation of men as Nobel Prize winners is a confirmation of the gender-science stereotype. That is, the bias that science is associated with men over women (Miller, Eagly, & Linn, 2014; Nosek et al., 2009). International research shows that even in countries with overall high gender equality (e.g., the Netherlands), men are more represented in the science fields. Consequently, people tend to evaluate the ability of men to perform science more positive than the ability of women to do so (Miller et al., 2014).

Although social stereotypes have been examined from different theoretical frameworks, the current research assumes stereotypes to occur due to associations between attributes and social groups (pseudocontingencies; PC's), as will be explained in the upcoming sections (Fiedler, Freytag, & Meiser, 2009; Kutzner & Fiedler, 2017; Meiser & Hewstone, 2010). The pseudocontingency literature shows that an initial bias in contrast with a stereotype can be induced, which can undo or reverse a stereotype (Fiedler, Freytag, & Unkelbach, 2007; Kutzner & Fiedler, 2017; Vogel, Kutzner, Fiedler, & Freytag, 2013). Interestingly, research by Harris, Fiedler, Marien, and Custers (2020) demonstrates that it is also possible for an initial bias to be upheld or attenuated, depending on certain conditions, as people continue with interacting. Nonetheless, little is known about whether a stereotype, in specific the gender-science stereotype, is attenuated or upheld when people continue with interacting. Therefore, the aim of the current research is to build on the research of Harris et al. (2020) by examining whether initial biases, regarding the gender-science stereotype are attenuated or upheld under certain conditions. It is proposed that initial biases will be attenuated when this bias will lead to negative outcomes. In contrary, it is expected that the initial biases will be upheld when this bias will lead to positive outcomes.

Social stereotyping

In the current research, stereotypes are looked upon from a cognitive-ecological perspective (Fiedler, 2000). Firstly, stereotypes are defined as subjectively expected statistical contingencies between attributes and social groups (Kutzner & Fiedler, 2017; Meiser & Hewstone, 2004). According to this cognitive-ecological approach, social stereotyping is dependent on the way people structure the information in their environment. Hence, social stereotypes originate in both the existing environmental structures and the adaptive strategies that individuals use to cope with their environment (Kutzner & Fiedler, 2017). Specifically, when organizing information about one's environment, individuals make use of mental shortcuts. Using mental shortcuts is a cognitive strategy in which part of the information is ignored, with the goal of making decisions more quickly, frugal and accurately than more complex methods (Gigerenzer & Gaissmaier, 2011). In an ideal world, individuals would make use of statistics when making decisions. However, in reality, where environments can be complicated and only part of the potential information is available, individuals rely on mental shortcuts in their decision-making process. One example of a complicated environment is a situation in which both the actions and the outcomes are skewed, making this environment more prone to mental shortcuts (Gigerenzer & Gaissmaier, 2011; Kutzner & Fiedler, 2017). Under such circumstances, people often form pseudocontingencies (PC's; Fiedler et al., 2007; Fiedler & Freytag, 2004). Making a PC inference reflects a categorical mistake, whereby an ecological correlation is mistaken for an individual contingency (Kutzner & Fiedler, 2017). In social stereotyping, information regarding individuals' attributes is assessed at an aggregate level, namely the social group to which individuals belong. Accordingly, this information is used to assess individuals' attributes (Fiedler et al., 2007). However, from an ecological correlation, one cannot logically infer an individuating correlation, therefore, representing a categorical error of stereotypical judgment (Kutzner & Fiedler, 2017). When looking at the gender-science stereotype, which indicates that science is associated with men over women (Miller et al., 2014; Nosek et al., 2009), an implicit ecological correlation can trigger a PC inference too, since the distribution of social groups and attributes is jointly skewed (Kutzner & Fiedler, 2017). In the science field, the social group 'men' is the more prevalent group, which will appear to be correlated with the more prevalent attribute 'ability to perform in mathematics'. Furthermore, the social group 'women' is the less prevalent group, which will appear to be correlated with the less prevalent attribute 'inability (or less ability) to perform in mathematics' (Kutzner & Fiedler, 2017; Meiser & Hewstone, 2010). People make these inferences, even though the relative rate of the prevalent attribute level is in fact equal in both groups, indicating that there is no actual difference in mathematical performance between men and women (Good, Aronsen, & Harder, 2008; Hargreaves, Homer, & Swinnerton, 2008; Reilly, Neumann, & Andrews, 2019). Therefore, the alignment of gender and ability to perform mathematics across ecologies yields an ecological correlation, only, while there does not need to be a contingency on an individual level (Kutzner & Fiedler, 2017).

As mentioned earlier, mental shortcuts are made in order to make the decision-making process more quickly, frugal or accurately (Gigerenzer & Gaissmaier, 2011). Additionally, PC inferences can afford efficient and adaptive strategies for individuals to deal with their limited cognitive capacity and uncertain environments (Fiedler et al., 2009; Kutzner & Fiedler, 2017; Meiser & Hewstone, 2010). Accordingly, it is often said that stereotypes contain a kernel of truth, namely stereotypes can be true at an ecological level, although not necessarily at an individual level (Kutzner & Fiedler, 2017). Therefore, PC's can lead to serious biases and flawed stereotypic expectations, which could result in significant harm (Fiedler et al., 2009; Kutzner & Fiedler, 2017). In particular, the PC inferences made in the gender-science stereotype can lead to negative consequences for individuals, especially the minority group, due to stereotype threat. Stereotype threat refers to the ramifications of an activated negative stereotype and emphasizes that members of a stigmatized group tend to perform worse on stereotype relevant tasks when confronted with that negative stereotype (Flore & Wicherts, 2015). Several meta-analyses support findings of stereotype threat in the gender-science stereotype, indicating that women score lower on mathematical tests when they are afraid to confirm to the gender-science stereotype (Flore & Wicherts, 2015; Nguyen & Ryan, 2008; Picho, Rodriguez, & Finnie, 2013; Walton & Cohen, 2003; Walton & Spencer, 2009). As a result, women's science career aspirations decrease and the distribution of men and women in the science career field remains skewed, suggesting a maintenance of the PC inference and the gender-science stereotype (Cundiff, Vescio, Loken, & Lo, 2013). Therefore, it is important to obtain insights in how stereotypes can be attenuated.

Stereotype change

Once a stereotype is established, it can function automatically and beyond awareness (Finnegan, Oakhill, & Garnham, 2015). Therefore, several studies state that stereotypes are impervious and pervasive and their influence is inevitable (Blair, 2002). Nonetheless, an overview of studies show that it is possible to change stereotypical beliefs through motivational, strategic and contextual influences (Blair, 2002). A method of stereotype change, which is shown to be an effective method in reducing stereotypic beliefs, is exposure to stereotype-inconsistent information (Dasgupta & Asgari, 2004; Dasgupta & Greenwald, 2001; Finnegan et al., 2015; Lai, Hoffman, & Nosek., 2013; Prati, Crips, & Rubini, 2015). This method implies that stereotypes can change by providing counterstereotypic associations of positive examples of disliked groups and negative examples or liked groups (Lai et al., 2013). Studies suggest counterstereotypic examples can lead to a reduction of race-bias and age-related prejudice for a duration of at least 24 hours after the intervention (Dasgupta & Greenwald, 2001). An example is the Obama effect, which refers to the impact of Barack Obama's presidential campaign in 2008. Because of his successful campaign, there was a high level of exposure to Obama, who provided a counterstereotypic black example, suggesting that African-Americans could also be well educated, motivated and articulate. This resulted in more positive attitudes of non-black people towards black people (Plant et al., 2009). More recent studies suggest this method to be effective in reducing other stereotypes, such as gender stereotypes (Dasgupta & Asgari, 2004; Finnegan et al., 2015; Prati et al., 2015). A specific example is the simulated classroom experiment of Fiedler et al. (2007), in which they show that it is possible to create a PC-inference contrary to the gender-science stereotype. Hence, this study showed that people are very sensitive to the base rates or the distribution they encounter (Fiedler et al., 2007). Interestingly, another study has demonstrated the effects of counterstereotypic examples of women (e.g., female leaders) in a real life setting (Dasgupta & Asgari, 2004). Specifically, this study compared two college campuses (i.e., a women's college and a coeducational college), in which the distribution of male and female faculty members varied. They showed that exposure to more female faculty members in a women's college resulted in a decrease of gender stereotypes one year later, compared to exposure to both male and female faculty members in a coeducational college. This effect was, however, mediated by the frequency with which participants were exposed to the counterstereotypic examples (Dasgupta & Asgari, 2004). In conclusion, exposure to counterstereotypic examples could be a valuable strategy for overcoming stereotypes (Dasgupta & Asgari, 2004; Dasgupta & Greenwald, 2001; Finnegan et al., 2015; Lai et al., 2013; Prati et al., 2015).

The exploration – exploitation tradeoff and different environments

Although people are aware that social stereotypes can be misleading in judgment, they keep relying on social stereotypes (Gigerenzer & Gaissmaier, 2011). This is due to the fundamental tradeoff between information search and reward maximization (e.g., the exploration – exploitation tradeoff), which all human beings face in life (Harris et al., 2020; Mehlhorn, et al., 2015). Because of exploration or information search, people gain more knowledge and learn about several possible options. However, when one option has been found to be superior, one can commit to this option and maximize the rewards of the chosen option, the most extreme form of exploitation (Harris et al., 2020). This fundamental tradeoff is relevant when processing stereotypes or stereotype-inconsistent information. When a stereotype is established, stereotype thinking becomes the default mode. More specific, the stereotype option has become superior and after commitment to this superior option, stereotypes start to guide judgment and behavior. By relying on these stereotypes, people save cognitive effort and time, since extensive information search is not necessary anymore. In other words, the stereotype option is exploited and its rewards are maximized (Crisp & Turner, 2011; Hutter, Crisp, Humphreys, Waters, & Moffit, 2009; Hutter & Crisp, 2005; Prati et al., 2015; Vasiljevic & Crisp, 2013). However, when stereotype-inconsistent information gets attention, such as in counterstereotypic examples, there is a shift out of the stereotypic thinking mode (Hutter, et al., 2009; Prati et al., 2015). Now, new options need to be explored, since the original match between social groups and attributes is not accurate anymore, which elicits a cognitive shift to more careful information processing in order to resolve the stereotypic inconsistency (Hutter & Crisp, 2005; Prati et al., 2015). Accordingly, this exploration of new options to recategorize the social groups and attributes, results in more individuating and non-stereotypical information (Prati et al., 2015).

Usually, balancing exploration and exploitation is essential for maximizing positive outcomes (Cohen, McClure, & Yu, 2007). Nonetheless, exploitation can result in negative outcomes, such as maintaining stereotypic beliefs (Prati et al., 2015). For example, Harris et al. (2020) show that the maintenance or attenuation of biases is not only dependent on the exploration – exploitation tradeoff, but also depends on the environment individuals find themselves in. Their research shows that in a reward-rich environment, in which positive outcomes are frequent, the initial bias is upheld. Specifically, since the outcomes are mainly positive, there is no need for further information search to receive positive outcomes, which triggers exploitative behavior (Cohen et al., 2007). On the other hand, in a reward-impoverished environment, in which most outcomes are negative, the initial bias is not rewarded, which results in an attenuation of the initial bias. More specifically, when outcomes remain negative, people remain searching for an effective strategy to receive positive outcomes, so the information search continues, which encourages explorative behavior (Cohen et al., 2007). Therefore, the environments influence whether either strategies of exploration or exploitation seem attractive or not (Harris et al., 2020).

The current research builds on the research by Harris et al. (2020), by applying these new insights to stereotype research. Therefore, the aim of the current research is to examine whether initial biases, regarding the gender-science stereotype, are attenuated or upheld in different environments. Specifically, it is proposed that initial gender-science biases will be exploited in an environment where this stereotype is rewarded. Accordingly, it is expected that this will result in the maintenance of the initial biases. In contrast, in an environment where this stereotype will lead to negative outcomes, it is expected that the initial biases will be attenuated, due to exploration of the available options. Eventually, the current research aims to get insight in the processes involved in the maintenance and attenuation of the gender-science stereotype, in order to provide insights that may help overcome damaging social stereotypes, such as the gender-science stereotype.

Methods

Participants

Based on an a priori power analysis, using G*Power, the sample size was estimated to be 179 participants (Faul, Erdfelder, Lang, & Buchner, 2007). These calculations were based on a 5% alpha-level, 80% statistical power and a medium effect size ($\eta = .25$). Participants were recruited via the online crowdsourcing platform Profilic Academic (https://profilic.ac/) and the experiment was run in English. Participants could participate for a financial reward varying between £0.85 and £1.15, depending on their performance. The initial sample consisted of $N = 218$ participants. After excluding participants who completed the experiment twice ($N = 6$) or did not fully complete the experiment ($N = 9$), the final sample consisted of $N = 197$ participants, of which $N = 99$ female, $N = 96$ male and $N = 2$ other gender participants. The ages of participants varied between 18 and 54 years old with an average of 29.58 ($SD = 9.96$) years old. Furthermore, 80,2% of the participants had an educational degree of College/A levels or higher. Moreover, the majority of participants did not have a psychology background (92.9%).

Design

This study used an experimental design, based on the experiment of Harris and colleagues (2020). In the current study, participants could repeatedly choose between a man and a woman to solve a mathematical problem with. Before each moment of choice, a mathematical problem is shown, with the question “who do you choose to solve this math problem with?”. The stimuli for the mathematical problems are derived from Washington State University (http://www.math.wsu.edu/math/HS/problems.html), which offers sample problems of intermediate algebra. For the stimuli of men and women, faces from the Radboud Faces Database are used (Langner, Dotsch, Bijlstra, Wigboldus, & Van Knippenberg, 2010). In this database several frontal images are validated, with respect to the shown facial expression, intensity and clarity of expression, genuineness of expression, attractiveness, and valence (Langner et al., 2010). A total of 14 faces (seven men and seven women) are selected from the database, based on a certain level of similarity and attractiveness. Participants are randomly assigned to conditions; which gender is shown frequently (i.e., men vs. women), which side the frequent gender is on (i.e., left vs. right), and overall experiencing either success or failure (i.e., reward-rich vs. reward-impoverished).

Procedure

The experiment is divided into four phases. Beforehand, to make the actual goal of the study less obvious, a situation is sketched. The participants are told that they take tutoring at school for math and two tutoring groups are available to help them, group A and group B, consisting of only men or only women. For each math problem participants need to solve, they can select a tutor from either of these groups. After the instructions, the experiment proceeds with an induction phase, in which an initial bias either in line with the original negative stereotype or in contrast with the original negative stereotype is induced. In all trials, participants are asked to answer the question “who do you choose to solve this math problem with?”. For the first 16 trials, participants are made to sample a particular distribution by only presenting one option on each outcome and controlling the feedback. Specifically, they were told that the computer would randomly determine which gender was to be chosen. After either a man or a woman who represents one of the tutor groups is chosen, the feedback would appear, that is, either success or failure in solving the math problem. Overall, in this phase, participants will either be winning or losing most of the time when guided to choose either a man or a woman. Since previous research has shown a distribution that can induce pseudocontingencies (cf. Experiment 2; Harris et al., 2020), the same distribution is used in the current study. Namely, of the 16 trials, participants experience nine times success when selecting a woman [man], and three times failure when selecting a man [woman].

After the initial phase, participants are asked to give estimates on their preferences regarding the tutoring groups (i.e., the genders). Firstly, they are asked with which gender they were more likely to succeed in solving the math problem, on which participants can answer by moving a slider which is anchored with the two genders displayed as images at the end. Furthermore, they are asked how likely it was to succeed (or fail) in solving a math problem if they chose women (or men), on which they can answer by moving a slider, anchored at 1% and 101%. In addition, participants are asked to indicate how confident they were in making a reasonable estimate regarding each gender, on which they can answer with a slider, anchored at "not confident at all" and "very confident". The variables were counterbalanced, such that participants are asked to give estimates regarding both genders and outcomes (e.g., success or failure).

The next phase is the free sampling phase of 84 trials, in which participants are free to choose either of the genders to solve a specific math problem with. Both genders are equal in their probabilities for positive outcomes, namely 75% of the time in the reward-rich condition and 25% in the reward-impoverished condition. The behavioral measure of this phase is the number of choices participants made for either of the genders (e.g., men or women).

At last, the relative preference estimates, conditional estimates, and confidence estimates are measured again. That is, participants were asked again to give estimates regarding the genders and outcomes, as they did after the initial phase.

Analysis

The results were analyzed using the *Statistical Program for Social Sciences (SPSS)*. Several analyses are conducted in order to test the hypotheses, among others one-sample and independent t-tests, and repeated measures Analysis of Variance (ANOVA).

Data preparation

Before the analyses, the dependent variables were counterbalanced, such that it did not matter which gender was demonstrated on the right or the left, which was frequent, and whether the participants either experienced a lot of success or failure (e.g. reward-rich vs. reward-impoverished condition). Furthermore, Δp-values (the difference between two conditional estimates; Allan, 1980) were calculated as a measure of perceived contingency between the frequent gender and the frequent outcome. Finally, the assumptions were checked, and all variables sufficiently met the assumptions required for the analyses conducted.

Results

Relative preference estimates pre-measure

After the induction phase, it was expected of participants to associate the frequent gender with the frequent outcome (success vs. failure), thus inducing an initial PC inference. In general, an initial bias is induced after the induction phase (M = 46.57, SD = 27.42, t(196) = -2.27, p = .024, d = -.16). Participants developed the same bias in both conditions, there was no difference between the reward-rich and reward-impoverished condition ($t(195) = 1.18$, $p = .238$, $d = -.17$). These results suggest that in both conditions, a preference for the frequent shown gender is created.

Conditional estimates pre-measure

As expected, after the initial phase, participants estimated that they had success with the frequent gender more often, than with the infrequent gender, again, suggesting a bias towards the frequent gender ($M = -.06$, $SD = .26$, $t(196) = -3{,}43$, $p = .001$, $d = -.24$). The mean Δp-score in the reward-rich ($\Delta p_{\text{rich}} = -.074$, $SD = .24$) and reward-impoverished ($\Delta p_{\text{impoverished}} = -.053$, $SD = .28$) did not differ from one another ($t(195) = .57$, $p = .573$, $d = -.08$), suggesting the bias is the same for both conditions. These results indicate that an initial PC inference is induced, linking the frequent gender with the frequent outcome, in both the reward-rich and the reward-impoverished condition.

Confidence pre-measure

Participants estimated their confidence similarly in the reward-rich ($M = -.03$, $SD = .24$) and the reward-impoverished condition ($M = .01$, $SD = .20$, $t(195) = 1.10$, $p = .271$, $d = -.16$).

Sampling

In contrast with the expectations, when comparing the sampling behavior with chance level, participants did not show a bias towards the frequent gender in sampling, ($M = .49$, $SD = .23$, $t(196) = -.83$, $p = .408$, $d = -.06$). Furthermore, there is no significant difference between conditions ($t(195) = .04$, $p = .972$, $d = -.01$). In both conditions, participants sampled the frequent option, on average 49% of the time ($SD_{\text{rich}} = .24$, $SD_{\text{impoverished}} = .22$). Furthermore, as can be seen at Figure 1., the pattern of sampling does not differ over trials. These results suggest the bias to be attenuated in both conditions, contrary the expectation that the bias should be upheld in the reward-rich condition.

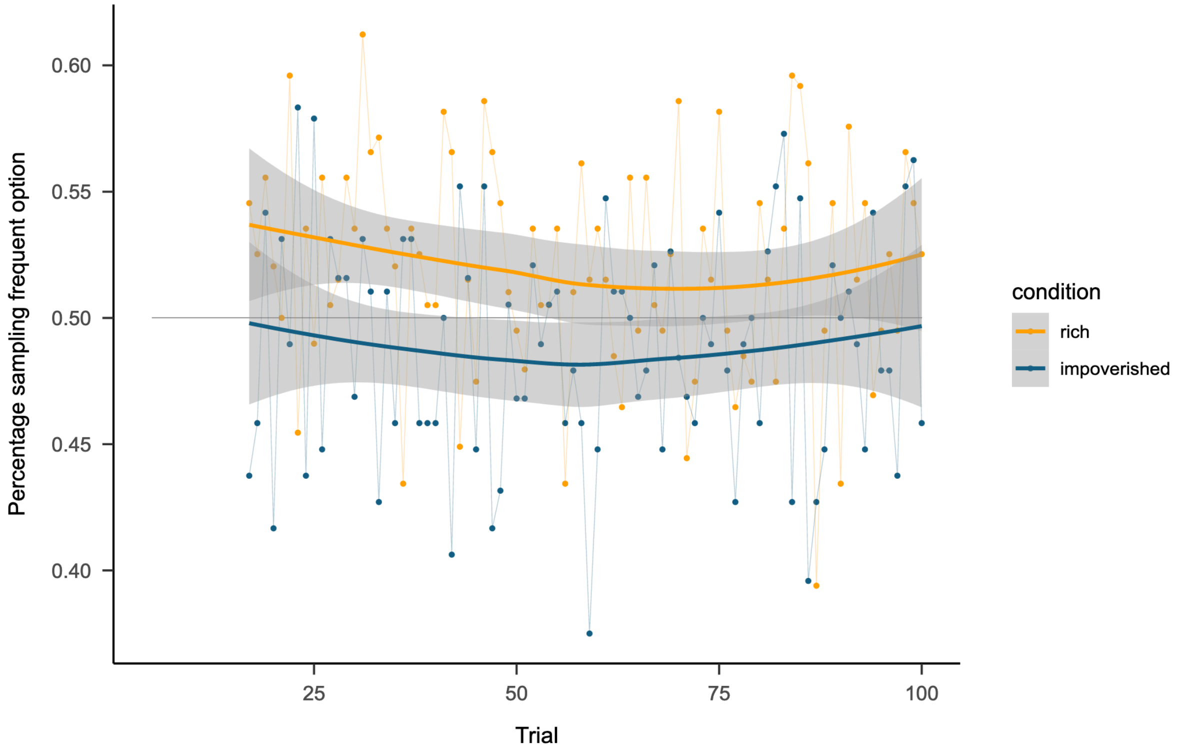

Figure 1. Percentage of participants sampling the frequent gender per trial. Chance level = 0.50.

Relative preference estimates post-measure

After the free sampling phase, it was expected for participants in the reward-rich condition to maintain the initial bias, but for participants in the reward-impoverished condition to attenuate the initial bias. In general, participants did not show any bias, when comparing the preference with chance level (M = 50.53, SD = 30.55, t(196) = -.22, p = .830, d = -.02). Furthermore, there are no differences between the reward-rich and the reward-impoverished condition (t(195) = 1.17, p = .245, d = -.17). These results, again, indicate that the bias is attenuated in both conditions, not only in the reward-impoverished condition as expected, but also in the reward-rich condition, in contrast with the expectations.

When comparing both measuring points, there is no difference between the pre-measure and the post-measure of the relative preference estimates (F(1, 195) = 2.62, p = .107, d = .21). Moreover, there is no effect of condition on this time difference (F(1, 195) = .01, p = .926, d = .01). Nonetheless, an interesting, although non-significant tendency can be seen. As demonstrated in Figure 2, participants in the reward-rich condition demonstrated initially more of a bias at the pre-measurement point, whilst participants in the reward-impoverished condition were already close to chance level (51) at that point. At the post-measuring point, the relative preferences of both conditions have been attenuated and moved towards chance level. However, the participants in the reward-rich condition, still show slightly more of the initial bias than the participants in the reward-impoverished condition. Also, it is shown that the participants in the reward-impoverished condition descriptively seem to overcompensate. Nonetheless, these tendencies are non-significant.

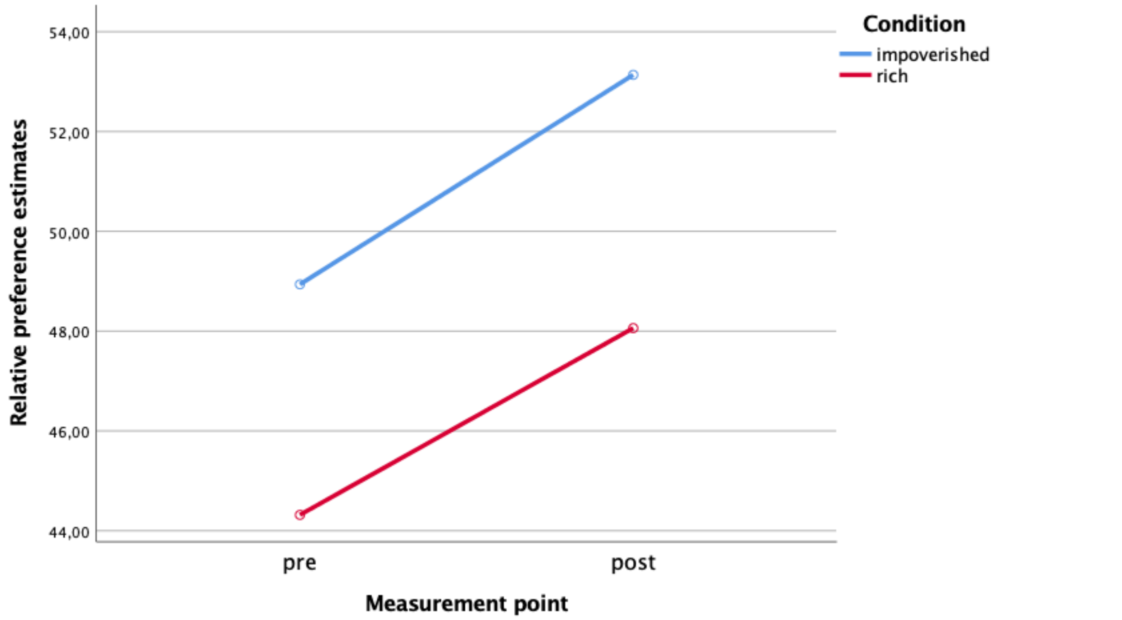

Figure 2. Relative preference estimates for the pre- and post-measurement for the frequent over the infrequent gender. Chance level = 51.00.

Conditional estimates post-measure

In contrast with the assumptions, participants estimated their chances of success with either of the options not significantly different ($M = -.03$, $SD = .33$, $t(196) = -1.22$, $p = .224$, $d = -.09$). Also, the mean Δp-score in the reward-rich ($\Delta p_{rich} = -.04$, $SD = .29$) and reward-impoverished condition ($\Delta p_{impoverished} = -.01$, $SD = .37$) did not differ from one another ($t(181,337) = .71$, $p = .480$, $d = -.10$). These results suggest that the frequent gender is no longer associated with the frequent outcome, thus, indicating that the initial bias is attenuated in both conditions.

When comparing both measuring points, there is no difference over time ($F(1, 195) = 1.88$, $p = .172$, $d = .20$). Furthermore, there is no effect of condition on time difference ($F(1, 195) = .06$, $p = .808$, $d = .04$). Interestingly, as with the relative preferences, the same non-significant tendency can be seen at the conditional estimates. As demonstrated in Figure 3., participants in the reward-rich condition show more of an initial bias at pre-measure, whilst the biases of the participants in the impoverished condition are located more closely towards chance level (.00) at the pre-measuring point. At the post-measuring point, the relative estimates are almost attenuated to chance level in the reward-impoverished condition. The conditional estimates of the reward-rich condition are also attenuated, but still remain more deviant from chance level. This tendency suggests that participants in the reward-rich condition still contain more of the initial bias than the reward-impoverished condition. Nonetheless, this tendency is non-significant.

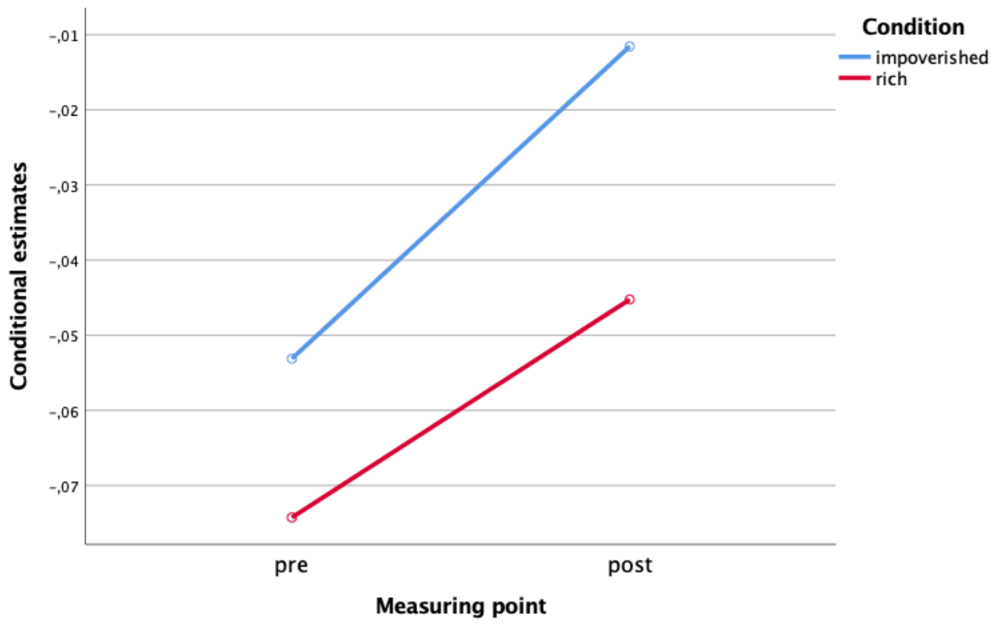

Figure 3. Δp-scores from conditional estimates for the pre- and post-measurement for the frequent over the infrequent gender. Chance level = .00.

Confidence post-measure

Finally, participants, again, did not estimate their confidence in their judgments to be different in the reward-rich (M = -.03, SD = .24) and in the reward-impoverished condition (M = -.01, SD = .34), t(169,992) = .63, p = .528, d = -.09).

Gender

Additional post hoc analyses were run to investigate the potential influence of gender of the tutors and gender of the participants. However, no significant differences have been found between tutors' gender, participants' gender or an interaction between both genders on all dependent variables (all $p > .054$).

Discussion

The aim of the current study was to test whether initial biases about gender groups would be maintained during an extensive sampling period. Specifically, the study examined whether initial biases regarding the gender-science stereotype, that is, the bias that science is associated with men over women (Miller et al., 2014; Nosek et al., 2009) were upheld or attenuated in different environments. Firstly, the current research shows that an initial PC inference, either in line with or in contrast with the gender-science stereotype can be created. Specifically, participants showed a preference for the frequent gender and linked the frequent gender with the frequent outcome, as expected from a PC inference. Afterwards, in the free sampling phase, as in Harris et al. (2020), it was assumed that the bias would be upheld in the reward-rich condition, but attenuated in the reward-impoverished condition. However, the current study does not support this hypothesis. The sampling patterns did not differ between the reward-rich and the reward-impoverished condition, and members from both the male and female tutoring group were chosen the same amount of time. Additionally, in contrast to the hypothesis, after the sampling phase, it was shown that the initial bias was attenuated in both conditions, suggesting that participants did not have a preference for either of the genders or linked the frequent gender with the frequent outcome. Also, no difference was found between the pre- and the post-measure. To summarize, the hypothesis that the initial bias would be upheld in the reward-rich condition, but attenuated in the reward-impoverished condition was not confirmed. Instead, the initial bias was attenuated in both conditions.

In the current study, a weak initial bias was found, but this bias was not strong enough to really affect the sampling. This finding can be explained by another process, that is set in motion when a bias is not strong enough, called subtyping (Caleo & Heilman, 2019; Richards & Hewstone, 2001). In the light of stereotypes, subtyping happens when atypical individuals are sectioned off from the overall bias and are seen as unrepresentative of the group as a whole (Richards & Hewstone, 2001; Weber & Crocker, 1983). Consequently, the bias is attenuated, and the situation returns to the status quo (Richards & Hewstone, 2001). It seems plausible that the process of subtyping has occurred in the current study. Specifically, during the sampling phase, participants could have started to individualize the tutors, thus, seeing them as belonging to neither the male or female tutoring groups, but as a subtype who was either successful or not successful. This process of individuating is explained by an exemplar-based model of social judgment, which states that specific past experiences with a target person or other individuals influence judgments and perceptions of people and groups (Smith & Zárate, 1990; Smith & Zárate, 1992). This theoretical model assumes that people possess many cognitive representations (i.e., exemplars), that include perceptual attributes, but also one's inferences, attributions and reactions. Accordingly, when a new instance is encountered, information from stored representations that is similar to this target is used to make judgments and inferences about this instance (Smith & Zárate, 1992). In the current study, this scenario could have occurred. Participants could have considered a tutor as successful or not successful, not per se because this tutor results in success or failure most of the time, but because some attributes of the tutor reminds the participant of their own experiences, inferences or reactions. In this instance, each tutor is judged individually, instead of being judged as tutoring group, resulting in a subtype of the tutoring group.

This process of individualization is supported by the empirical notion that this process is more likely in a smaller sample of new instances (Caleo & Heilman, 2019; Meiser & Hewstone, 2006; Weber & Crocker, 1983). The current study used seven faces per tutor group, and none of the tutors resulted in success or failure all the time. Additionally, in the sampling phase, first all seven faces were shown before participants would see one of them again, so they rarely encountered someone back to back. Moreover, studies suggest that engaging in exemplar-based categorization is more likely when people do not receive prior information about the category representatives, in this case the tutor groups (Smith & Zárate, 1990). In the current study, little prior information was given regarding the two tutoring groups and it is not very likely that people have strong prior opinions about tutoring groups. Accordingly, their categorization had to be based on participants' past experiences, self-attributes, group memberships and the social context, contributing to exemplar-based categorization (Smith & Zárate, 1990).

Furthermore, a more specific issue with participants' past experiences that could have led to exemplar-based categorizing, might be participants' prior biases regarding gender-stereotypes in general or the gender-science stereotype in specific. Because these stereotypes are quite pervasive, it is likely that participants possessed preexisting biases (Blair, 2002; Finnegan et al., 2015). Accordingly, empirical evidence suggests that high-prejudiced individuals are more likely to engage in subtyping, whereas low-prejudiced individuals are less likely to do so (Riek, Mania, & Gaertner, 2013). More concrete, when participants encounter the new tutor, some attributes of this tutor might be similar to their preexisting biases, which makes them categorize these tutors alongside their exemplar representations, instead of categorizing them in one of the tutoring groups. Hence, the initial bias, that might have been just strong enough after its induction, is attenuated during the intensive sampling period, due to individual judgment of each tutor, resulting in subtyping. Additionally, empirical evidence suggests that the more cognitive load people experience, the more they start to rely on past experiences, existing knowledge and stereotypes (Gigerenzer & Gaissmaier, 2011; Spears, Haslam, & Jansen, 1999). It is conceivable that the 84-trial sampling period could increase the cognitive load that the participants experience, making exemplar-based categorizing more likely. Concluding, since the initial bias was weak, throughout the course of the experiment, tutors were evaluated individually and subtypes were formed. Consequently, the initial biases were attenuated in both the reward-rich and the reward-impoverished condition. This explanation seems plausible, since an interesting tendency can be seen, indicating that in the reward-rich condition the bias is somewhat stronger than in the reward-impoverished condition, but the difference is just not strong enough. In the current study, we did not test participants' past experiences, inferences, attributions, reactions or preexisting biases, so we cannot rule out their influence. Since it is proclaimed that these factors could act as possible mediators or moderators (Riek et al., 2013; Smith & Zárate, 1992), we recommend future research to include participants' exemplar representations in the experiment. This could be established by adding an Implicit Association Test (IAT) at the start of the experiment. Accordingly, the results of the IAT can be taken into account, and can be compared with the measures of the experiment to test the influence of prior biases as a possible moderator or mediator.

Another explanation for the non-significant results after the sampling phase, is that the process of maintaining or attenuating a bias, in specific regarding the gender-science stereotype, is complex. Several other effects, could have affected the induction and maintenance of the bias by allowing too much noise in the research design. First, the study design is based on the experiment of Harris and colleagues (2020), in which they used 16 trials to induce a bias towards a brown bag, labeled A and B with wins or losses. In this design, the bags are neutral stimuli, whereas in the current study faces of male and female tutors are used to induce this bias. Faces are not neutral, as they contain lots of social information, and people extract all kinds of information from seemingly small attributes of the face (Calder & Young, 2005; Rule et al., 2008). Thus, even though we attempted to rule these effects out by using faces from the Radboud Faces Database (Langner et al., 2010), it seems impossible to avoid the effects of other appearance characteristics, such as attractiveness. Perhaps, 16 trials were sufficient to induce a bias that was powerful enough to be maintained regarding a neutral brown bag. However, in order to induce and maintain a bias regarding human faces, more trials might be necessary to induce a more lasting bias. We recommend future research to take this into account, for example, by increasing the amount of trials in the induction phase. Secondly, in the current study we induced a PC-inference regarding the skewness of frequent and infrequent attributes and outcomes. Nonetheless, the social world is skewed in many more ways, among others in pleasant vs. unpleasant stimuli, interesting vs. uninteresting knowledge domains, relevant vs. irrelevant to a task focus, normative vs. non-normative behavior and ingroups vs. outgroups. In the current study, these other skews could be of influence on the PC-inference we tried to induce (Kutzner & Fiedler, 2017). As mentioned earlier, even though we tried to use validated stimuli, faces contain lots of social information, resulting in participants interpreting some stimuli as more pleasant or less pleasant (Calder & Young, 2005; Rule et al., 2008). In sum, these alternative effects could result in too much noise in the current study. Therefore, the current research design might be too simplistic to obtain a close look into the effects of the maintenance and attenuation of initial biases. We recommend future research to carefully examine these other effects in order rule out their influence, and hopefully get a more precise insight into how stereotypes can be changed.

Notwithstanding the limitations, the current study has contributed to this field of research, since the study was first in broadening the findings of Harris and colleagues (2020) to a stereotype context, specifically, the gender-science stereotype. The study tried to extend and encounter some interesting findings, such as the induction of the initial bias in a context of gender and math ability. However, the initial bias did not affect sampling, which could be the result of a non-sufficient induction of the bias, or the result of several unknown moderators or mediators. Nonetheless, these insights are interesting, because they show it takes more than just a few trials and a skewed distribution to induce a group-wide bias, at least when the group is something as artificial as a tutoring group. Furthermore, the research implies some important moderators and mediators involved in stereotype change to be tested in future research, such as exemplar representatives. This could be done by improving the research design, for example by adding an Implicit Association Test. At last, this research is based on the assumption that biases can be maintained, since exploitation limits the information that we learn about other choice alternatives. We tried to obtain insights on how to accomplish stereotype change by providing choice alternatives to the gender-science stereotype, such as female tutors that are successful or male tutors that are unsuccessful. Even though we did not fully obtain change, we believe that an easy way to overcome this exploitation, resulting in maintenance of biases, is to sometimes force the sampling of other choice alternatives. In addition to the scientific relevance, this study has some societal implications as well. Specifically, the current study recommends companies in the science field to show variations to the initial gender-science stereotype, such that women are associated with the ability to perform sciences too. Accordingly, more and more women can be stimulated to strive for careers in the science field, with the promising hope of a fourth female Nobel Prize winner in the future.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*, 147-149. doi:10.3758/BF03334492

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and social psychology review*, *6*, 242-261. doi:10.1207/S15327957PSPR0603_8

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641–651. doi:10.1038/nrn1724

Caleo, S., & Heilman, M. E. (2019). What could go wrong? Some unintended consequences of gender bias interventions. *Archives of Scientific Psychology*, *7*, 71–80. doi:10.1037/arc0000063

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 933–942. doi:10.1098/rstb.2007.2098

Crisp, R. J., & Turner, R. N. (2011). Cognitive Adaptation to the Experience of Social and Cultural Diversity. *Psychological Bulletin*, *137*, 242–266. doi:10.1037/a0021840

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender-science stereotypes predict science identification and science career aspirations among undergraduate science majors? *Social Psychology of Education*, *16*, 541–554. doi:10.1007/s11218-013-9232-8

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658. doi:10.1016/j.jesp.2004.02.003

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814. doi:10.1037/0022-3514.81.5.800

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*, 659–676. doi:10.1037/0033-295X.107.4.659

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology*, *87*, 453–467. doi:10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An Integrative Account of an Intriguing Cognitive Illusion. *Psychological Review*, *116*, 187–206. doi:10.1037/a0014480

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudocontingencies in a Simulated Classroom. *Journal of Personality and Social Psychology*, *92*, 665–677. doi:10.1037/0022-3514.92.4.665

Finnegan, E., Oakhill, J., & Garnham, A. (2015). Counter-stereotypical pictures as a strategy for overcoming spontaneous gender stereotypes. *Frontiers in Psychology*, *6*, 1–15. doi:10.3389/fpsyg.2015.01291

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*, 25–44. doi:10.1016/j.jsp.2014.10.002

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*, 451–482. doi:10.1146/annurev-psych-120709-145346

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, *29*, 17–28. doi:10.1016/j.appdev.2007.10.004

Hargreaves, M., Homer, M., & Swinnerton, B. (2008). A comparison of performance and attitudes in mathematics amongst the 'gifted'. Are boys better at mathematics or do they just think they are? *Assessment in Education: Principles, Policy and Practice*, *15*, 19–38. doi:10.1080/09695940701876037

Harris, C., Fiedler, K., Marien, H., & Custers, R. (2020). Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments. *Journal of Experimental Psychology: General,* Publish Ahead of Print. doi:10.1037/xge0000754

Hutter, R. R. C., & Crisp, R. J. (2005). The composition of category conjunctions. *Personality and Social Psychology Bulletin*, *31*, 647–657. doi:10.1177/0146167204271575

Hutter, R. R. C., Crisp, R. J., Humphreys, G. W., Waters, G. M., & Moffitt, G. (2009). The dynamics of category conjunctions. *Group Processes and Intergroup Relations*, *12*, 673–686. doi:10.1177/1368430209337471

Kutzner, F., & Fiedler, K. (2017). Stereotypes as pseudocontingencies. *European Review of Social Psychology*, *28*, 1–49. doi:10.1080/10463283.2016.1260238

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing Implicit Prejudice. *Social and Personality Psychology Compass*, *7*, 315–330. doi:10.1111/spc3.12023

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, *24*, 1377–1388. doi:10.1080/02699930903485076

Lunnemann, P., Jensen, M. H., & Jauffred, L. (2019). Gender bias in Nobel prizes. *Palgrave Communications*, *5*. 17–20. doi:10.1057/s41599-019-0256-3

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215. doi:10.1037/dec0000033

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, *87*, 599–614. doi:10.1037/0022-3514.87.5.599

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, *36*, 315–336. doi:10.1002/ejsp.304

Meiser, T., & Hewstone, M. (2010). Contingency learning and stereotype formation: Illusory and spurious correlations revisited. *European Review of Social Psychology*, *21*, 285–331. doi:10.1080/10463283.2010.543308

Miller, D. I., Eagly, A. H., & Linn, M. C. (2014). Journal of Educational Psychology Women ' s Representation in Science Predicts National Gender-Science Stereotypes : Evidence From 66 Nations Women ' s Representation in Science Predicts National Gender-Science Stereotypes : Evidence From 66 Nations. *Journal of Educational Psychology, 107,* 631–644. doi:10.1037/edu0000005

Nguyen, H. H. D., & Ryan, A. M. (2008). Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence. *Journal of Applied Psychology*, *93*, 1314–1334. doi:10.1037/a0012702

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., Somogyio, M., Akramid, N., Ekehammard, B., Vianellop, M., Banaji, M.R., Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10593–10597. doi:10.1073/pnas.0809921106

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *Journal of Social Psychology*, *153*, 299–333. doi:10.1080/00224545.2012.737380

Plant, E. A., Devine, P. G., Cox, W. T. L., Columb, C., Miller, S. L., Goplen, J., & Peruche, B. M. (2009). The Obama effect: Decreasing implicit prejudice and stereotyping. *Journal of Experimental Social Psychology*, *45*, 961–964. doi:10.1016/j.jesp.2009.04.018

Prati, F., Crisp, R. J., & Rubini, M. (2015). Counter-stereotypes reduce emotional intergroup bias by eliciting surprise in the face of unexpected category combinations. *Journal of Experimental Social Psychology*, *61*, 31–43. doi:10.1016/j.jesp.2015.06.004

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*, *49*, 25–50. doi:10.1007/s11165-017-9630-6

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review*, *5*(1), 52–73. https://doi.org/10.1207/S15327957PSPR0501_4

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2013). Reverse Subtyping: The Effects of Prejudice Level on the Subtyping of Counterstereotypic Outgroup Members. *Basic and Applied Social Psychology*, *35*, 409–417. doi:10.1080/01973533.2013.823616

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and Awareness in the Perception and Categorization of Male Sexual Orientation. *Journal of Personality and Social Psychology*, *95*, 1019–1028. doi:10.1037/a0013194

Smith, E. R., & Zárate, M. A. (1990). Exemplar and Prototype Use in Social Categorization. *Social Cognition*, *8*, 243–262. doi:10.1521/soco.1990.8.3.243

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. In *Psychological Review, 99,* 3–21. doi:10.1037//0033-295x.99.1.3

Spears, R., Haslam, S. A., & Jansen, R. (1999). The effect of cognitive load on social categorization in the category confusion paradigm. *European Journal of Social Psychology*, *29*, 621–639.

doi:10.1002/(SICI)1099-0992(199908/09)29:5/6<621::AID-EJSP969>3.0.CO;2-W

Vasiljevic, M., & Crisp, R. J. (2013). Tolerance by Surprise: Evidence for a Generalized Reduction in Prejudice and Increased Egalitarianism through Novel Category Combination. *PLoS ONE*, *8*. doi:10.1371/journal.pone.0057106

Vogel, T., Kutzner, F., Fiedler, K., & Freytag, P. (2013). How majority members become associated with rare attributes: Ecological correlations in stereotype formation. *Social Cognition*, *31*, 427–442. doi:10.1521/soco_2012_1002

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Walton, G. M., & Spencer, S. J. (2009). Intellectual Ability of Negatively Stereotyped Students. *Psychological Science*, *20*, 1132–1139. doi:10.1111/j.1467-9280.2009.02417.x

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, *45*, 961–977.

doi:10.1037/0022-3514.45.5.961