# University of Groningen

## The boomerang effect of psychological interventions

Levy, Aharon; Maaravi, Yossi

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

Link to publication in University of Groningen/UMCG research database
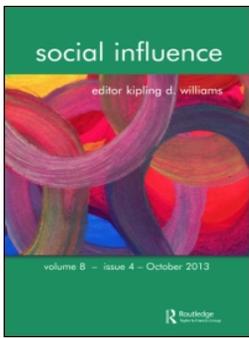
# The boomerang effect of psychological interventions

Aharon Levy & Yossi Maaravi

Published online: 28 Dec 2017.

Submit your article to this journal ☐

Article views: 1965

View Crossmark data ☐

Routledge
Taylor & Francis Group

# The boomerang effect of psychological interventions

Aharon Levy[a,b] and Yossi Maaravi[c]

[a]The Ivcher School of Psychology, Interdisciplinary Center, Herzliya, Israel; [b]Department of Social Psychology, University of Groningen, Groningen, The Netherlands; [c]The Adelson School of Entrepreneurship, Interdisciplinary Center, Herzliya, Israel

**ABSTRACT**
Research has found that teaching people about psychological biases can help counteract biased behavior. On the other hand, due to the innate need for preservation of a positive self-image, it is likely that teaching people about biases they hold, may cause a boomerang effect in cases where being associated with a specific bias implies negative social connotations. In the three studies below we examine situations in which psychological bias implies negatively associated behavior, and show that teaching people about bias in those contexts can be counterproductive.

## Introduction

Past research in behavioral decision-making has long shown that people's judgments and decisions are often biased or irrational (Kahneman, 2011). This is the case even when these decisions are made by experts (Loschelder, Friese, Schaerer, & Galinsky, 2016; Northcraft & Neale, 1987), or in cardinal situations such as legal verdicts (Danziger, Levav, & Avnaim-Pesso, 2011; Englich, Mussweiler, & Strack, 2006) or presidential elections (Antonakis & Dalgas, 2009; Ballew & Todorov, 2007; Todorov, Mandisodza, Goren, & Hall, 2005). Interestingly, one method that can be used to counteract such biases is simply raising the awareness of people to the bias and teaching them its effects. For example, research has found that simply teaching participants about Naïve realism – i.e. the psychological bias that causes people to perceive information that does not correspond with their existing perspective as inherently wrong – caused those participants to become more open to new and different ideas (Nasie, Bar-Tal, Pliskin, Nahhas, & Halperin, 2014).

On the other hand, research on the psychology of self-affirmation suggests that using the above debiasing technique may cause a boomerang effect in cases where being associated with the specific bias that a person allegedly holds also implies negative social connotations (Schumann & Dweck, 2014; Sherman & Cohen, 2006; Steele, 1988). Moreover, the work done in the field of commitment and *post hoc* justification has found that negative consequences of an action can increase the likelihood that a person will justify it and resist change (Axsom & Cooper, 1985; Frenkel & Doob, 1976; Kiesler,1971). Both of these fields of research can be seen as derived from the cognitive dissonance theory in that when an

---

**CONTACT** Aharon Levy ✉ a.l.levy@rug.nl

individual is faced with information that does not coincide with his world view (i.e. positive self-image), he is likely to find a way to discredit the new information in the attempt resolve the dissonance that is created by the said information (Festinger, 1962).

Accordingly, it is possible that an attempt to change peoples' biased behavior in a manner that threatens their self-image, or challenges their world view will result in a negative boomerang effect. This boomerang effect may also result in entrenchment, that is, not only that they will not change their judgments or behavior, but they will even fortify their existing opinions. Think, for example, of a moderate conservatives during the 2016 presidential elections campaign who are considering to vote for Trump. Once they are exposed to information regarding racist and misogynistic biases that may be affecting Trump supporters, are they likely to accept the fact that they too might be affected by those biases, and reconsider their political behavior, or are they more likely to backlash after being implicitly (or explicitly) accused of being misogynistic racists? If the latter option takes place, it is also likely to assume that this backlash will cause them to be more inclined to justify their actions using other explanations, and in turn reinforce their original judgment and behavior. Thus, building on existing research, we suggest that before moving on with the development of psychological interventions that are based on exposure to psychological bias, it is important to examine this possible boomerang effect. Additionally, the examination of peoples' reaction to information regarding psychologically biased behavior can add an unexplored domain to the commitment and *post hoc* justification fields of research.

When considering the conditions under which exposure to psychological bias might backlash there are two main factors that need to be considered. First, the bias itself can have inherent negative association. For example a bias against powerful women can be considered a socially problematic trait by society due to its inherent chauvinistic nature. On the other hand, a bias related to inaccuracy in judging character based on specific traits does not necessarily have a negative social connotation. The second factor to take into consideration in terms of the boomerang effect is the social context in which the bias is described. For example, a bias that makes you prefer one White male over another White male might lead to inaccuracy, but does not have any negative social connotations. On the other hand, the same bias that will make you prefer a male over a female might be perceived as deplorable, as a result of the misogynistic outcome. What we are suggesting here can be seen as a 2×2 matrix in which if the bias does not have inherent socially sensitive nature, and the context is not socially sensitive then exposure to the existence of the bias is likely to have the positive effect described in the literature (Nasie et al., 2014). However, if the bias is inherently socially sensitive, or presented in a socially sensitive context (or both) it is likely to expect a negative boomerang reaction to the accusation of being biased (see Table 1).

**Table 1.** Possible combinations of bias nature and bias context. We hypothesize that only combination #1 will enable reducing the bias by raising the awareness to it.

|  | Inherently neutral bias | Inherently sensitive bias |
|---|---|---|
| Neutral social context | Exposure to the bias diminishes the impact of the bias (#1) | Exposure to the bias causes boomerang effect and increases the impact of the bias (#2) |
| Sensitive social context | Exposure to the bias causes boomerang effect and increases the impact of the bias (#3) | Exposure to the bias causes boomerang effect and increases the impact of the bias (#4) |

Based on this 2×2 matrix we designed a pilot and three experimental studies that examine the different possible combinations. In the pilot study we tested several different biases and contexts to make sure that they are indeed perceived as socially sensitive by participants. In study 1 we replicated the results of Nasie et al. (2014), and demonstrated how teaching people about a neutral bias that might be relevant to their behavior in a neutral context changed their behavior and reduced the bias (combination #1). In Study 2 we taught decision makers about an inherently sensitive bias that may imply they were holding chauvinistic views in a sensitive social context, of female representation in politics (combination #4). Finally, to test if a sensitive social context is enough to evoke the boomerang effect even toward a neutral bias, in Study 3 we taught decision makers about a neutral bias in a sensitive social context of a gender in the work place (combination #3). In Studies 2 & 3, we hypothesized that the boomerang effect predicted by self-affirmation literature (Schumann & Dweck, 2014; Sherman & Cohen, 2006; Steele, 1988) would be a result of participants being motivated to justify their actions and stay set in their ways in order to avoid being labeled as chauvinistic or misogynistic. Since inherently sensitive biases are always linked to sensitive social contexts combination #2 was covered by Study 2 as well, and did not call for another Study.

## Pilot study

In order to verify the social sensitivity of the biases and context we were going to examine in the experimental studies we ran a pilot study in which we asked participants to assess the perceived social sensitivity of different biases and different contexts on a scale from 1 to 10. We recruited 153 participants via Mturk (92 male; $M_{age}$ = 35.04 years, SD = 9.9). The biases we examined were: 1. The 'powerful women bias' in which powerful women are perceived in a negative light and judged more harshly than men performing in the same manner. We expected this bias to be perceived as inherently socially sensitive due to its chauvinistic nature. 2. The 'Halo effect bias' in which a specific evaluation of an individual influences evaluations of other attributes of that person. We expected this bias to be perceived as inherently socially neutral. 3. The 'track 1 track 2 bias' in which people make decisions in a quick and instinctive manner even when a more analytical approach is preferable. We expected this bias to be perceived as inherently socially neutral as well. The two contexts we examined were: A context in which two men from the same race and social class are competing over a given position as the neutral context, and a context in which a man and a woman are competing over a given position as the sensitive context.

As expected the powerful women bias was perceived as significantly more socially sensitive ($M$ = 6.08) than both the Halo effect bias ($M$ = 3.93, $t$ = 8.86, $p$ < .001), and the track 1 bias ($M$ = 4.43, $t$ = 6.45, $p$ < .001). Additionally, the context in which a man was competing with a women was perceived as significantly more sensitive ($M$ = 6.36) than the context in which a man was competing with a man (2.96, $t$ = 14.96, $p$ < .001). Moreover, in order to verify that we are indeed dealing with biases that are socially sensitive and biases that are not, as opposed to biases that are all sensitive just to different degrees we performed a one sample t test, and compared the ratings of the different biases and contexts to the mid point of the scale (5.5). Once again corroborating our hypothesis, the sensitive bias and context were significantly higher than the scale mid point (powerful women bias $t$ = 31.09, $p$ < .001; man vs. woman context: $t$ = 29.29, $p$ < .001). On the other hand, the neutral biases and

context were significantly lower than the scale mid point (halo effect: $t = 17.94$, $p < .001$; track 1: $t = 20.94$, $p < .001$; man vs. man: $t = 11.65$, $p < .001$). These findings verified our hypothesis regarding the perceived social sensitivity of the different biases and contexts we chose to examine in the following experimental studies.

## Study 1

In study 1 we attempted to replicate the results of Nasie et al. (2014) in a scenario in which the bias and context at hand are not socially sensitive, and thus raising awareness to the bias can reduce its effect. Specifically in Study 1 we focused on political voting behavior, and examined if teaching actual voters about a bias, that might be relevant to their voting but does not reflect poorly on their self-image, changed their behavior and reduced the bias. We ran Study 1 in the context of the 2015 elections in Israel. As we describe below, the campaigns during these elections put a significant emphasis on the candidates' physical attributes. This discourse, regarding the physical attributes of both candidates, and their relevance to political decision-making, can be explained by the aforementioned research on the *Halo effect*. The Halo effect describes a phenomenon in which a specific evaluation of an individual influences evaluations of other attributes of that person (Nisbett & Wilson, 1977; Thorndike, 1920). This effect has been found to be especially prominent in the case of physical attributes and political competence, where rapid judgments about the person- ality traits of political candidates are made based solely on the physical appearance of the said candidates (Todorov et al., 2005; Verhulst, Lodge, & Lavine, 2010; and see Olivola & Todorov, 2010 for a review)

In view of that, the following study was an attempt to inform voters about this Halo effect bias, in order to raise their awareness, and have them focus on crucial political subjects or more relevant leadership characteristics instead of physical appearance. Additionally, since the implications of being affected by this specific bias would only cause someone to choose one White male candidate over another, being affected by this bias would not entail a racist or chauvinistic implication and would therefore not cause a boomerang effect. Accordingly, our hypothesis was that similar to the findings of Nasie et al. (2014) mentioned in the introduction, it would be enough to simply teach voters about the existence of the Halo effect bias in order to counteract it. In order to test this hypothesis we designed and ran the following experiment.

## Method

### *Participants and design*

Eighty-one participants (50 male; $M_{age} = 38.20$ years, SD = 13.55) were recruited via an inter- net survey company, that also ran the election polls for one of the two main news channels in Israel (Chanel 10) . The participants, who were randomly selected from a nationwide panel, participated in the study and filled out a questionnaire, all on the Qualtrics platform. They were randomly assigned to either the control condition or the experimental condition. We determined the sample size based on the findings in the Nasie et al. paper (2014) that found a medium size effect ($d = .46$). Accordingly, we ran a power analysis (through G*Power, Faul, Erdfelder, Lang, & Buchner, 2007) that assumed we wanted to be able to achieve a

statistical power of .80 to detect a medium-sized effect ($d$ = .5). This analysis suggested a required sample size of 37 participants per condition. The experiment was conducted on March 9th 2015 only one week before the elections and in the peak of the electoral process.

### Procedure

In the 2015 elections in Israel the incumbent prime minister had a clear advantage over the opposition leader. While the prime minister Bejamin Netanyahu was considered to have exceptional charisma and an outstanding presence, the opposition leader Isaac Herzog was referred to as lacking any charisma due to his high pitched voice, and a feminine facial structure (Black, 2015). Indeed, even Herzog's campaign managers were dragged into this debate of their candidate's problematic physical characteristics and launched a campaign in which he appeared with a deep manly voice-over and presented his credentials. At the end of the commercial he then posed a question: 'After all you've heard you're still not going to elect me because of my voice?' (Jerusalem post, 2015).The participants were all given a short article to read. The article was supposedly from the science section of a newspaper and described research which found that people have psychological biases that cause them to link external attributes, such as facial structure, or height, to leadership abilities of congress candidates. The article also stated that this is despite the fact that there is no real correlation between these elements. In order to avoid demand characteristics the article described a study conducted in the US, and charisma and pitch of voice which were the main issues in the Israeli campaign, were not mentioned in the article. In the control condition the article was about astronomy. After reading the article and answering a few basic comprehension questions, the participants filled out a survey asking them how much both leading candidates for prime minister, Netanyahu and Herzog, were fit to be prime minister on a scale of 0–100.

### Results

As expected, in the control condition Netanyahu, the more charismatic, deep voiced, manly candidate, was found to be significantly more fit to be prime minister ($M$ = 65.78, SD = 34.26 vs. $M$ = 39.87, SD = 34.41), $t$ = 2.74, $p$ = .009, $d$ = .75. On the other hand, in the experimental condition where participants were exposed to the information regarding the psychological bias of attributing leadership skills to external attributes, the differences between the two leaders was almost completely eliminated leaving Netanyahu with a negligible advantage ($M$ = 52.19, SD = 32.04) over Herzog ($M$ = 47.39, SD = 32.05), $t$ = .51, $p$ = .61, $d$ = .15, see Figure 1.

These results verified our hypothesis by both diminishing the charismatic leader's status as well as enhancing the less charismatic opponent's status. Interestingly, a deeper analysis of the results found that most of the change in the perception of the candidates probably stemmed from the participants who displayed the more extreme positions to begin with. We used chi-square tests to analyze the proportions of participants who rated either Netanyahu or Herzog less than 5 or more than 95 in each of the two experimental conditions. Our results indicated that while in the control condition 29% of the participants extremely favored Netanyahu, after being exposed to the existence of the psychological bias in the
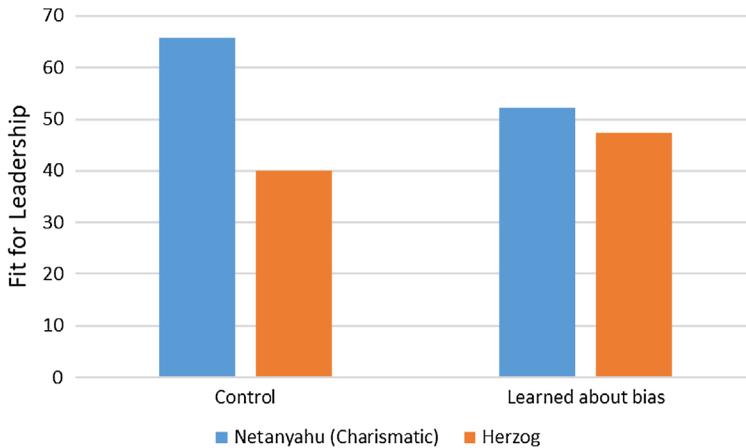
**Figure 1.** Difference between candidates in control condition vs. learning about bias condition.

experimental condition, only 8% of participants showed such extreme favoritism: $X^2$ (2, $N = 81) = 5.33$, $p < .05$. We did not find similar effects in the three other categories.

## Discussion

As expected, the results of Study 1 replicated the findings of Nasie et al. (2014), and raising the awareness of the participants to the psychological bias that was affecting their political judgment led to a de-biasing process that diminished the support for the physically superior candidate, and enhanced the support for the physically inferior candidate. Additionally, the replication of this effect in a real world political context increased the external validity of these findings. However, based on the hypothesis presented in the introduction, we assumed that the successful de-biasing occurred due to the fact that there were no negative social implications for being influenced by the Halo effect. On the other hand, when dealing with a socially sensitive bias or context, a boomerang effect may take place in which not only does the awareness to an alleged bias not reduce that bias but it even enforces it, as a result of people trying to maintain a positive self-image in light of prejudice accusations. In Study 2 we tested this possible boomerang effect in the context of male chauvinism with a socially sensitive bias in a socially sensitive context.

## Study 2

In order to test the effects of bias awareness raising efforts with a socially sensitive bias in a socially sensitive context, we replicated the design of Study 1 only this time we chose a bias that may entail a socially negative connotation, namely: chauvinism in the socially sensitive context of female representation in politics. In Study 2 instead of asking about the how fit the party leaders were for the prime minister position, we asked about the fitness of the political parties themselves to lead the government taking onto consideration that there was a significant gender difference between the two parties. The two leading parties in the said elections were the Likud party, and the Labor party. While the Likud party only had 1 woman in the top 10 party candidates, the Labor party had 4 women in the party's top 10,

and also ran on a feminist and gender equality platform. Accordingly, the bias we chose to raise awareness to was the powerful women bias in which women who behave confidently and assertively are not as well received as men who engage in the same behaviors (Rudman, 1998). We hypothesized that raised bias awareness would cause a boomerang effect with male voters. That is, men, who are at risk of being perceived as chauvinists if they harbor such a bias, would not react positively to the bias awareness and perhaps even backlash. On the other hand, women are less at risk of being perceived as misogynist thanks to being women (indeed, research has found that people are less affected by derogatory labels used by their group to describe their group, Carnaghi & Maass, 2007). Accordingly, we also hypothesize that raised bias awareness should have a positive effect on women and promote more egalitarian behavior among them similar to the findings in Study 1.

## Method

### Participants and design

One hundred and eleven participants (62 male; $M_{age}$ = 40.77 years, SD = 14.91) were recruited via the same internet survey company as in Study 1. The participants, who were randomly selected from a nationwide panel, participated in the study and filled out a questionnaire, all on the Qualtrics platform. They were randomly assigned to either the control condition or the experimental condition. We based our decision to determine the sample size on a power analysis (through G*Power, Faul et al., 2007) that assumed we wanted to be able to achieve a statistical power of .80 to detect a medium-sized effect ($d$ = .5) but this time with four groups due to the expected gender moderation. This analysis suggested a required sample size of 54 participants per condition. This experiment was conducted on March 15th 2015 only two days before the elections and in the peak of the electoral process.

### Procedure

The participants were all given a short newspaper article to read. The article was supposedly from the science section of a newspaper and described research which found that people have psychological biases that cause them perceive powerful women in a negative light and judge them more harshly than men in the same position performing in the same manner. In order to avoid demand characteristics, the article did not directly address the elections or the political parties at hand. In the control condition the article was about astronomy. After reading the article and answering a few basic comprehension questions, the participants filled out a survey asking them how much both leading political parties, Likud and Labor, were fit to be in the government on a scale of 0–100.

## Results

As expected, men reacted negatively to the allegations of harboring chauvinistic biases and in the experimental condition men rated the dominantly male Likud party even higher than they did in the control condition ($M$ = 61.72, SD = 33.39 vs. $M$ = 44.84, SD = 34.28), $t$ = 1.96, $p$ = .05, $d$ = .50. Additionally, men rated the gender equal Labor party marginally lower in the experimental bias awareness condition than they did in the control condition

($M$ = 52.41, SD = 38.23 vs. $M$ = 68.79, SD = 29.13), $t$ = −1.91, $p$ = .06, $d$ = −.48. Moreover, when comparing the difference between the two parties in each condition the same pattern appears where in the control condition the labor party is rated as more fit to govern than the Likud party and in the experimental condition it is revered, and the Likud party is deemed more fit to govern ($M$ = 23.93, SD = 54.71 vs. $M$ = −9.31, SD = 63.63), $t$ = −2.21, $p$ = .03, $d$ = .56. We did not find any significant differences in term of the female participants in the Likud rating ($p$ = .71) the Labor rating ($p$ = .53), or the difference between parties ($p$ = .89) see Figure 2.

## Discussion

As expected, not only were we not able to replicate the positive effects of raising bias aware-ness as in Study 1, but male participants in Study 2 that were exposed to information regarding a bias than could portray them in a negative social light, reacted 'negatively' to the bias awareness manipulation. Accordingly, the intervention that was intended to promote gender equality ended up harming the political party that ran on gender equality, and assisted the party that lacked in this aspect. Nonetheless, study 2 has a few limitations. First, despite the fact that the alleged newspaper article did not address political elections directly, it did explicitly describe the bias against dominant/successful women. Thus, it is possible that this influenced the answers of the participants in the experimental condition. Second, it is possible that political affiliation of participants may have played a significant role in participants' response in Studies 1 and 2. Despite the random assignment of partic-ipants addresses this issue, it is still a variable we cannot completely account for. This issue could account for the gender based difference between the control condition and for the lack of difference between conditions among female participants. Finally, while running the study adjacent to the elections gave the findings of studies 1 and 2 a boost of relevance, participants were probably bombarded with political campaigns during the same period without our ability to properly control for other political stimuli. Study 3 addressed these three main limitations.
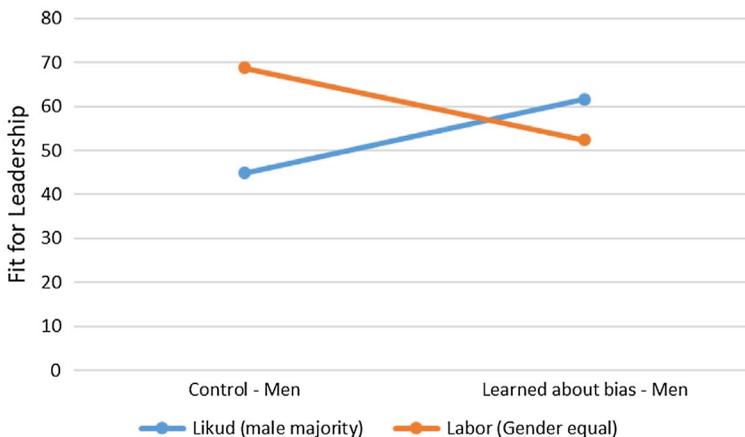


**Figure 2.** Difference between rating of parties among male participants in the control condition vs. learning about bias condition.

## Study 3

Study 3 was designed in a workplace context where political affiliation would not act as a confound, and where there is seemingly no explanation for discrimination against women other than gender-based bias. As in the previous studies, Study 3 attempted to raise awareness to a bias affecting decision-making and to test whether such raised awareness would have a productive or a counterproductive effect in the attempt to reduce biased behavior. However, in Study 3 we examined a scenario in which only the context is socially sensitive while the bias itself is neutral in order to examine whether this was enough to induce the boomerang effect. Once again, we hypothesized that men – who are at risk of being perceived as chauvinists if they harbor a bias that effects their behavior towards women – would backlash and react negatively to the bias awareness intervention. One the other hand, women that are in less risk of being perceived as misogynists should have a positive reaction to the bias awareness intervention, and display more egalitarian behavior in the gender context as in Study 2.

## Method

### Participants and design

One hundred and twenty-one participants (66 male; $M_{age}$ = 32.54 years, SD = 14.91) were recruited via Mturk. The participants, who were randomly selected, participated in the study and filled out a questionnaire, all on the Qualtrics platform. They were randomly assigned to either the control condition or the experimental condition. The sample size was rationale was similar to the one in Study 2.

### Procedure

The participants were all given a short newspaper article to read. For the experimental group, the article was supposedly from the science section of a newspaper and described research regarding decision-making. In order to avoid demand characteristics, the article did not directly address gender but instead described research which found that people have two possible decision-making processes: track 1 and track 2. While track 1 is more fast and intuitive, track 2 is more analytical and information based. Although track 1 is more instinctive, it is prone to mistakes and exposed to biases and therefore it is advised to try and be more analytical and patient when in decision-making processes in order to avoid biased decisions. Thus, participants who may negatively biased against women were not explicitly accused of that. Rather, the text implied that they use track 1 in their judgements and consequently may judge women more harshly. In the control condition participants read an article about Zebras.

After reading the article and answering a few basic comprehension questions, the participants were told that Mturk is considering the employment of an Mturk worker to act as workers' representative and manage worker rights issues etc. It said that there were two possible candidates for the job, and Mturk would like to get the workers opinion on them based on their bio. The structure of the two bios was identical and consisted of personal information, education and grades, and past experience, as well as some neutral filler information. The main difference between the bios was that one was a bio of a man – Kevin, and the other was the bio of a woman – Natalie.

To control for physical attractiveness (Mobius & Rosenblat, 2006) and other facial characteristics that may influence participants' judgements (Todorov, Said, Engell, & Oosterhof, 2008), we used the faces by Gladstone and O'Connor (2014), used a computer program to create similar faces of a man and a woman (see Figure 3).

Importantly, and to counter the possible alternative explanation that Kevin was a better candidate for the job, Natalie was a better fit for the job on almost every front. After reading the bios participants were asked to rate the extent to which they felt each candidate was fit for the job (job-fit) and the extent to which they felt the participants had the necessary skills for the job (job-skills) on a scale of 0–100.

## Results

The two above variables (job-fit and job-skills) were integrated to a single grade (Kevin: $r = .81$, $p < .001$; Natalie: $r = .69$, $p < .001$). To examine our hypothesis that gender would moderate the reaction to the bias awareness manipulation, we ran a moderation analysis (using Preacher and Hayes' PROCESS macro, model 1; Hayes, 2013) on the rating of the candidates, considering participants condition (with/without bias awareness), and participants' gender as independent variables. As expected, the analysis yielded a significant two-way interaction between the bias awareness manipulation and gender, $b = 11.29$, $SE = 5.94$, $t = 1.90$, $p = .06$. Although not surprisingly all participant rated the far superior Natalie more positively than Kevin (based on the given information), among female participants the difference between Natalie and Kevin's rating in the control condition was only 6.65 (SD = 14.47), however, in the bias awareness condition it grew to 11.55 (SD = 12.21) in Natalie's favor ($p = .18$, $d = .37$). On the other hand, the male participants showed the exact opposite trend and although they favored Natalie by 12.50 points (SD = 17.65) in the control condition, after being exposed to their possible bias the difference between Natalie and Kevin Shrunk to 6.10 points (SD = 17.14) ($p = .16$, $d = .37$) See Figure 4.



**Figure 3.** The similar faces of Kevin and Natalie. Source: Gladstone and O'Connor (2014).
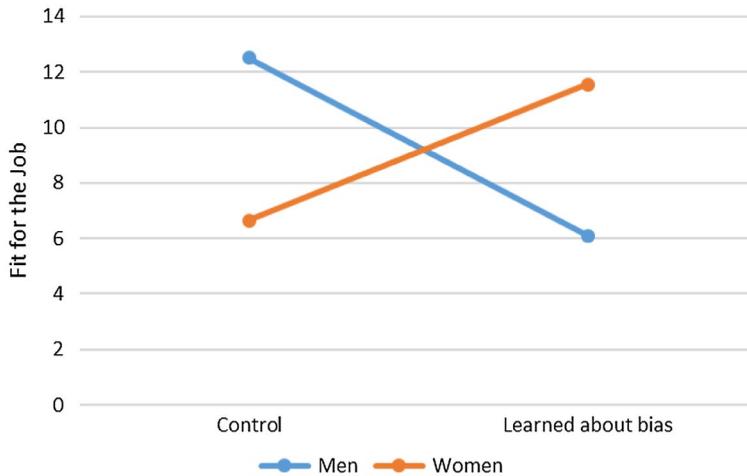
**Figure 4.** Difference in favor of the more suitable female candidate in the control condition vs. learning about bias condition.

## Discussion

Study 3 replicated the results of Study 2 even without explicitly arising awareness to the sensitive social topic, and this time in a more controlled environment without any political ideology confound. As in Study 2, the results of Study 3 show that participants exposed to a psychological bias they may hold, which suggested negative social connotations such as chauvinism reacted in a reverse way: increasing rather than decreasing their biased judgments.

## Conclusion

While previous work done by Nasie et al. (2014), which has also been replicated in Study 1 above, has shown that teaching people about psychological bias can help counteract this bias, in the current article we offer a more complex process. Based on the finding in Studies 2 and 3, it seems that in situations where being affected by psychological bias can entail negative social branding such as racism or chauvinism, people taught about that psychological bias might actually react in the opposite manner than expected and become more entrenched in their biased behavior. Additionally, based on the findings from Study 3 it is apparent that if the context is socially sensitive then raising awareness to psychological bias can backfire even when the bias at hand is neutral in nature.

Returning to the 2016 US elections example, we suggest that the above mechanism might have been at work there too. If the democrats thought that many American voters were over- focusing on the private email-server issue, thus suffering from the prominence effect (Tversky, Sattath, & Slovic, 1988), raising their awareness to this bias might have worked as a de-biasing tool (Nasie et al., 2014). But not only that the democrats did not use this technique, but they kept stressing the alleged chauvinistic or racist agenda of their adversary, thus implicitly accusing potential Trump voters of holding such socially negative opinions, and possibly making them even more set in their ways.

While studies 2 and 3 supported our hypothesis, the underlying mechanism at hand still needs to be brought to light. Based on the self-affirmation literature (Schumann & Dweck, 2014; Sherman & Cohen, 2006; Steele, 1988), we assume that in order to maintain a positive self-image, people exposed to information regarding their own biased behavior might be motivated to justify their actions in alternate ways and as a result also fortify their existing biased behavior. Future studies should try and examine this issue, and flesh out the underlying mechanism.

In sum, although teaching people about their psychological bias can be beneficial in counteracting biased behavior, it seems that it can also backfire under certain circumstance. To this end, we tried to provide an initial framework for a more complex approach to psychological bias information as a social intervention in the attempts to devise a positive and useful socio-psychological intervention.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323*(5918), 1183–1183.

Axsom, D., & Cooper, J. (1985). Cognitive dissonance and psychotherapy: The role of effort justification in inducing weight loss. *Journal of Experimental Social Psychology, 21*(2), 149–160.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*(46), 17948–17953.

Black, I. (2015). Isaac Herzog, scion of Israeli political dynasty, battling to dislodge Netanyahu. *The Guardian*. http://www.theguardian.com/world/2015/mar/17/isaac-yitzhak-herzog-profile-israeli-political-dynasty-zionist-union

Carnaghi, A., & Maass, A. (2007). In-group and out-group perspectives in the use of derogatory group labels: Gay versus fag. *Journal of Language and Social Psychology, 26*(2), 142–156. Chicago.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences, 108*(17), 6889–6892.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, *32*(2), 188–200.

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.

Festinger, L. (1962). Cognitive dissonance. *Scientific American., 207*(4), 93–106.

Frenkel, O. J., & Doob, A. N. (1976). Post-decision dissonance at the polling booth. *Canadian Journal of Behavioral Science, 8*(4), 347.

Garramone, G. M. (1985). Effects of negative political advertising: The roles of sponsor and rebuttal. *Journal of Broadcasting & Electronic Media, 29*(2), 147–159.

Gladstone, E., & O'Connor, K. M. (2014). A counterpart's feminine face signals cooperativeness and encourages negotiators to compete. *Organizational Behavior and Human Decision Processes, 125*, 18–25.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kiesler, C. A. (1971). *The psychology of commitment: Experiments linking behavior to belief*. San Diego, CA: Academic Press.

Loschelder, D. D., Friese, M., Schaerer, M., & Galinsky, A. D. (2016). The too-much-precision effect: when and why precise anchors backfire with experts. *Psychological science*, *27*(12), 1573–1587.

Mobius, M. M., & Rosenblat, T. S. (2006). Why beauty matters. *The American Economic Review, 96*(1), 222–235.

Nasie, M., Bar-Tal, D., Pliskin, R., Nahhas, E., & Halperin, E. (2014). Overcoming the barrier of narrative adherence in conflicts through awareness of the psychological bias of naive realism. *Personality and Social Psychology Bulletin*, 0146167214551153.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*(1), 84–97.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 ms: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83–110.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology, 74*(3), 629.

Schumann, K., & Dweck, C. S. (2014). Who accepts responsibility for their transgressions? *Personality and Social Psychology Bulletin*, 0146167214552789

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in experimental social psychology, 38*, 183–242.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology, 21*, 261–302.

The Jerusalem Post staff. (2015). Herzog's new ad chides voters unsure of his voice. *The Jerusalem Post*. http://www.jpost.com/Israel-Elections/Watch-Herzogs-new-dubbed-ad-chides-voters-unsure-of-his-voice-391271

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*(1), 25–29.

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

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological review*, *95*(3), 371.

Verhulst, B., Lodge, M., & Lavine, H. (2010). The attractiveness halo: Why some candidates are perceived more favorably than others. *Journal of Nonverbal Behavior, 34*(2), 111–117.