

Social Psychological Bulletin

Psychologia Społeczna

Are Women Perceived as Worse Magicians Than Men? Gender Bias When Evaluating Magic Tricks

Pascal Gygax^a, Cyril Thomas^b, André Didierjean^c, Gustav Kuhn^d

[a] *University of Fribourg, Fribourg, Switzerland.* [b] *University of Paris, Paris, France.* [c] *University of Franche-Comté, Besançon, France.* [d] *Goldsmith University, London, United Kingdom.*

Social Psychological Bulletin, 2019, Vol. 14(3), Article e33574, <https://doi.org/10.32872/spb.v14i3.33574>

Received: 2019-02-01 • Accepted: 2019-09-19 • Published (VoR): 2019-11-13



Handling Editor: Natasza Kosakowska-Berezecka, University of Gdansk, Gdansk, Poland

Corresponding Author: Pascal Gygax, University of Fribourg, Avenue de l'Europe 20, 1700 Fribourg, Switzerland.
E-mail: pascal.gygax@unifr.ch

Supplementary Materials: Data, Materials [see [Index of Supplementary Materials](#)]



Abstract

We present two experiments investigating the effect of the perceived gender of a magician on the perception of the quality of magic tricks. In Experiment 1, tricks performed by an allegedly female magician were considered worse than those by an allegedly male magician. In Experiment 2, participants had to generate possible solutions to how the tricks were done. Under these conditions, male participants were better at explaining the tricks, but the gender effect found in Experiment 1 disappeared. We discuss the gender bias in Experiment 1 and the lack of bias in Experiment 2 in terms of specific social and cognitive mechanisms (e.g., cognitive dissonance).

Keywords

magic, social biases, stereotypes

In the seminal paper by [Goldberg \(1968\)](#), college women were more critical when evaluating journal articles that appeared to be written by a woman (i.e., Joan T. McKay) than when written by a man (i.e., John T. McKay). Although the statistics presented at the time were rather elementary, the effects appeared larger for articles written in male-dominated fields (i.e., law and city planning) than neutral (i.e., linguistics and art history) or female-dominated topics (i.e., elementary school teaching and dietetics). [Goldberg \(1968\)](#) interpreted this gender bias as illustrating the distortion that precedes prejudice. Gener-



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

ally, these results illustrated how gender can act as a catalyst for biased representations of reality.

Although this seminal paper strongly indicated prejudicial beliefs against women (and by women), some authors have questioned its actual relevance (e.g., Swim, Borgida, Maruyama, & Myers, 1989). Aside from its obvious statistical weakness (i.e., Goldberg only presented descriptive statistics), Swim et al. (1989), in their meta-analysis, argued that subsequent studies that addressed such a gender bias were characterized by weak effect sizes. Swim et al. (1989) did acknowledge, though, that gender bias effects could emerge when people have less information about the protagonists. For example, only giving a protagonist a name might be sufficient to activate the bias. However, when more information about the protagonist is provided, the bias diminishes.

Although the underlying mechanisms modulating this gender effect are still not fully understood (see Koch, D’Mello, & Sackett, 2015, for a discussion on possible mechanisms), some factors that enhance the effect have been identified. First, stereotypical expectations play an important role in activating gender biases. Second, the effect seems to emerge more strongly when stereotypically male-dominated activities are evaluated (e.g., Koch et al., 2015; Colley, North, & Hargreaves, 2003; Proudfoot, Kay, & Koval, 2015; Swim et al., 1989).

In this paper, we concentrate on *magic*, a domain that has received little attention in terms of these mechanisms, yet it is strongly male-dominated, or at least considered as such. In a recent norming study, Misersky et al. (2014) collected social norms for more than 420 occupations and activities across 7 languages (i.e., Czech, English, French, German, Italian, Norwegian, and Slovak), and found that the work of a *magician* was considered as more likely to be carried out by men. On average, across all languages, participants considered that only 26% of magicians were women (29% in the French-speaking part of Switzerland where the present study was carried out). This perceived proportion of women as magicians has been stable over the last decade (i.e., 24% in the French part of Switzerland in 2008; Gabriel, Gyga, Sarasin, Garnham, & Oakhill, 2008).

This male-magician prevalence could be anchored in the history of magic, where women have essentially been relegated to “lovely assistants” (Bruns & Zompetti, 2014). It has also often been difficult for women to gain access to this rather secretive organization, and for many years women were denied access to major magic associations (Nardi, 1988). For example, London’s Magic Circle – one of the largest magic associations – granted permission for women to enter the association only in 1991. From a historical perspective, women performing magic throughout the 14th, 15th, and 16th centuries were often associated with witchcraft. In fact, women behaving in any non-normative ways were often associated with witchcraft, and subsequently burnt or punished (Chollet, 2018) and during the early 19th century, many found success only in roles as psychics or mediums (Bruns & Zompetti, 2014; Nardi, 1988). Houdini’s later embodiment of masculinity set the standards in the magic entertainment industry (Mangan, 2007), which con-

fined women to being *weak* and *vulnerable* (Bruns & Zompetti, 2014), or *powerless* (Nardi, 1988).

We would argue that these attributes – and their founding history – constitute a perfect example of *gender stereotypes*. Gender stereotypes can broadly be defined as generalized beliefs and expectations about social roles or occupations, that are considered appropriate based on individuals' socially identified sex (Eagly, 1987). These roles can often be summarized along two main dimensions: communion and agency (e.g., Eagly, 1987, Eagly, 1997; Eagly & Wood, 1991). Whereas communal attributes refer to *friendliness*, *unselfishness*, *concern for others* and *emotion expressiveness*, and are associated with women, agentic roles refer to *independence*, *assertiveness*, *instrumental competence* and *masterfulness*, and are associated with men (Eagly & Wood, 1991). Attributes typically associated with magicians, such as *power* and *control* (Nardi, 1988), clearly fall under agency. Through different socialization processes, these attributes become internalized as part of an individual's self-concept and personality, and act as basic foundations for subsequent behaviors (Eagly, 1997). In turn, these stereotypes influence how we evaluate other people's behaviors (Eagly & Wood, 1991). For example, Colley et al. (2003) asked participants to listen to instrumental music extracts composed by allegedly female or male composers. Afterwards, they were required to evaluate the extracts on several musical competence items. The authors argued that since music composition has historically been considered a stereotypically male-dominated activity (Green, 1997), participants should rate them more highly when the piece is introduced as composed by a man (i.e., Klaus Behne or Simon Healy) than by a woman (i.e., Helena Behne or Sarah Healy). Participants did indeed give lower ratings on adjectives relating to musical competence for the female pieces. This effect was strongly alleviated when the name was accompanied by the composer's alleged biography. Similarly, Proudfoot et al. (2015) found that for an identical architectural outcome, men were evaluated as more creative than women.

Others have found similar gender biases when focusing on male-dominated activities, even when short descriptions of the targeted protagonists were provided. For example, Szesny, Spreemann, and Stahlberg (2006) showed participants short descriptions of female or male protagonists, together with a picture and a name tag (i.e., Mrs or Mr Keller). The short descriptions contained leadership characteristics, which are stereotypically male. When asked to recall the leadership characteristics, participants felt more certain that the characteristics were present when these were associated with a man than when associated with a woman. Evaluations have also been shown to be biased when more subtle gender cues were given. For example, Fleischmann, Sieverding, Hespeneide, Weiß, and Koch (2016) showed that women wearing a feminine outfit (a dress and some makeup) – presented in a picture – were judged as having fewer computer skills than those with a neutral outfit (trousers and no makeup). Significantly, the success of a woman with a neutral outfit was more likely to be attributed to *skill* compared to that of a woman wearing a feminine outfit, where success was attributed to *luck*. This is in line

with the seminal work by Deaux and Emswiler (1974) who showed that women performing well in a stereotypically masculine task (i.e., a perceptual discrimination task of mechanical objects) were considered to be “lucky”, whereas men were considered to be “skilled”. Of importance – at least when considering the expectancy-value theory of achievement motivation (Eccles & Wigfield, 2002) – different attributions of success lead to different career, educational or activity choices. As such, investigating gender stereotypes in masculine contexts, such as magic, may uncover important factors that explain the relatively low prevalence of women in these contexts.

The Present Experiments

The aims of the present experiments were twofold. First, we wanted to investigate the extent to which magic – one of the most male dominated art forms (Misersky et al., 2014) – activates gender attributes that are potentially linked to judgment biases. To address this, we followed Goldberg’s (1968) procedure and presented magic tricks as either performed by a woman (i.e., Nathalie) or by a man (i.e., Nicolas) (Experiment 1). We expected that magic tricks allegedly performed by a woman would be evaluated more negatively than when presented as performed by a man (Hypothesis 1).

Second, in Experiment 2, we expected that participants would evaluate a magic trick more positively when not (really) knowing how it was done. More specifically, we predicted that evaluating a trick negatively whilst at the same time not knowing how it was done would generate tension, or *cognitive dissonance* (Festinger, 1957). To reduce this dissonance, one would most likely evaluate it more positively. To test this idea, the second experiment used the same information about the magician (i.e., their name), but asked participants to engage in an additional task: generate possible solutions. The gender cues therefore remained constant across experiments. This manipulation is rather different to those used in previous gender bias studies. Previous studies simply added additional information about the protagonists, which seemed to alleviate the gender bias (e.g., Colley et al., 2003; and as discussed by Swim et al., 1989).

Accordingly, in Experiment 1, we expected participants to evaluate tricks performed by Nathalie more negatively than for Nicolas, but this difference should disappear in Experiment 2 (Hypothesis 2). In sum, if people struggle to come up with the true explanation of the tricks, they should evaluate them (especially those by Nathalie) more positively.

Experiment 1

Method

Participants

Sixty-four psychology undergraduate students from the University of Fribourg took part in this experiment (*Mean age* = 22.80; *SD* = 4.15; *n* = 33 women). All students were granted course credit for participation and were part of a first-year research method class. All participants had granted written informed consent before the experiment.

Design and Procedure

Participants were presented with fourteen video sequences, each presenting a close-up (i.e., close proximity) magic trick. The magicians wore white gloves and a white long-sleeve t-shirt, which prevented participants from identifying cues about the magician's true gender (hereafter called *sex of the magician*). The magic tricks were performed by a female and male magician and all were performed in front of a neutral background. Each participant saw seven sequences by the female magician, and seven by the male one. At the end of the experiment, participants were asked whether they noticed that there were two different magicians, and none of them did. Two lists were created, to ensure that each trick was performed by both magicians across the experiment, and that each trick was only presented once per participant.

For each participant, the magic tricks were presented in a random order. After each magic trick, participants had to answer three questions: (1) *How good was the trick?* (1 = *not good at all* to 7 = *very good*), (2) *How impressive was the trick?* (1 = *not impressive at all* to 7 = *very impressive*), (3) *Did you guess how the trick was performed?* (1 = *not at all* to 7 = *yes, I am sure of it*). The former two questions were aimed at providing us with crucial quality evaluations. The first question assessed a more global evaluation, and the second question assessed an evaluation more specific to magic. In essence, a trick considered as good may not necessarily be seen as impressive (this is especially true for tricks that one believes to have guessed how they were done¹), and we wanted to make sure we covered all possible evaluative impressions. The latter question (i.e., guessing question), consequently, explored whether the trick evaluations were impacted by whether participants discovered (or thought they had discovered) the solutions to the tricks.

Before starting the experiment, participants were presented with a cover story informing them that they had to evaluate some magic tricks. They were told that the tricks would be presented to the public, but that we wanted to get insight into how people appreciate different types of tricks. Half of the participants were told that the tricks were performed by NATHALIE (i.e., female magician), and half were told they were performed

1) For example, knowing how David Copperfield made the Statue of Liberty disappear most likely affects the impressiveness of the trick (i.e., it is less impressive), yet it remains a very good and ingenious trick.

by NICOLAS (i.e., male magician). We chose a between-participant design to avoid participants discovering the purpose of the experiment (i.e., manipulating gender). Both names are very frequent in the French part of Switzerland, with the former unambiguously referring to a woman and the latter to a man. To ensure that participants would not forget who was performing the tricks, we repeated the name of the magician before each video sequence, by specifying: “Are you ready for the next trick by [NAME]”. Participants simply pushed the space bar to start the video. Videos were between 3 to 21 seconds long (Male magician: $M = 10.4$, $SD = 5.84$; Female magician: $M = 10.2$, $SD = 5.77$).

Results

In order to include both participants and items (i.e., magic tricks) as random factors in all analyses and to avoid any fixed effect fallacies by separating by-participant and by-item analyses (Brybaert, 2007; Clark, 1973), data were analyzed by fitting linear mixed-effects models using the R software (R Development Core Team, 2010, version 3.1.2). Models were tested using the *lmer()* function of the *lmer4* package of R, and model comparisons were assessed using the *anova()* function, which calculates the Chi-square value of the log-likelihood in order to evaluate the difference between models, following Baayen’s (2008) procedure. Models were compared using a forward-testing approach, from the simplest model to more complex ones, as advocated by Field (2014), and as commonly done in psycholinguistics research (e.g., Öttl & Behne, 2016). Namely, fixed effects (main and interaction) were included one at a time, and each resulting model was compared to a model that did not include the added factor. As justified by the design and to the extent that the model still converged, we included Sex of Magician as a random slope for items. The reason for doing so was that although changing the magician within each session (i.e., for each participant) remained unnoticed by participants, the quality of the magic tricks was not completely homogeneous, as shown by an overall slightly better – yet non-significant – evaluation of our male magician (male magician: $M = 4.21$, $SD = 0.84$; female magician: $M = 3.96$, $SD = 1.03$, $t(13) = 1.28$, $p = .22$, 95% CI_{difference} [-0.67, 0.17]) (our female magician was less experienced, as indicated by a lower number of years of practice compared to our male one). Finally, to obtain *p-values* for our final model, we used the *summary()* function from the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017)².

Results for the Question “How Good Was the Trick?”

When comparing our random model – only encompassing items and participants as random factors (and Sex of Magician as random slope for items), to one also including Perceived Gender of Magician (NATHALIE vs. NICOLAS), the latter showed a better fit,

2) The original data and analysis R scripts (of Experiments 1 & 2) can be accessed at <https://osf.io/tqxs5/>

$\Delta\chi^2 = 5.80$, $\Delta df = 1$, $p = .02$. However, adding the Respondent's Sex or Sex of Magician did not improve the model.

The final model (see Table 1 and Figure 1), including only Perceived Gender of Magician as a fixed factor, showed that participants rated the tricks performed by NICOLAS more positively, $M = 4.20$; $SD = 0.95$, 95% CI [3.89, 4.52], than those performed by NATHALIE, $M = 3.96$; $SD = 1.10$, 95% CI [3.57, 4.34], $t(170.9) = 2.36$, $p = .019$. It is interesting to note that numerically, Nathalie is just below the mid-point, whereas Nicolas is just above it.

Table 1

Model Estimates for Our Best Fitting Model of How Good a Trick Was

Factor	Estimate	SE	df	t	Pr(> t)
(Intercept)	4.36	0.27	25.99	16.37	< .001
Perceived gender of magician (female)	-0.45	0.19	178.33	-2.42	.017

Note. Includes Gender of Magician as a fixed factor, participants and items as random intercepts, and Sex of Magician as random slope for items. Treatment contrasts were used.

Results for the Question “How Impressive Was the Trick?”

The correlation between the first and second measure (*How good is the trick?* and *How impressive is it?*) was very high, $r(62) = .94$, 95% CI [.90, .96], $p < .001$. Not surprisingly then, when comparing our random model – which only encompassed items and participants as random factors (with both random intercept and random slope) –, to one also including Perceived Gender of Magician (NATHALIE vs. NICOLAS), the latter showed a better fit, $\Delta\chi^2 = 5.23$, $\Delta df = 1$, $p = .02$. However, adding the Respondent's Sex or Sex of Magician did not improve the model.

The final model (see Table 2), including only Perceived Gender of Magician as a fixed factor, showed that participants rated the tricks performed by NICOLAS more positively, $M = 4.11$, $SD = 0.89$, 95% CI [3.79, 4.43], than those performed by NATHALIE, $M = 3.88$, $SD = 1.01$, 95% CI [3.51, 4.24], $t(161) = 2.15$, $p = .033$. Again, numerically, Nathalie is just below the mid-point, whereas Nicolas is just above it.

Table 2

Model Estimates for Our Best Fitting Model of How Impressive a Trick Was

Factor	Estimate	SE	df	t	Pr(> t)
(Intercept)	4.25	0.28	24.15	15.33	< .001
Perceived gender of magician (female)	-0.43	0.18	170.08	-2.29	.023

Note. Includes Gender of Magician as a fixed factor, participants and items as random intercepts, and Sex of Magician as random slope for items. Treatment contrasts were used.

Results for the Question “Did You Guess the Magic Trick?”

The correlation between the first and third question (*How good is the trick?* and *Did you guess how the trick was performed?*) was lower than between the first and second question, and was negative, $r(62) = -.49$, 95% CI $[-.66, -.28]$, $p < .001$. The less participants thought they understood how the trick was performed, the more likely a trick was to be evaluated better.

The initial random model (as for the previous question) was neither improved by the Gender of Magician, $\Delta\chi^2 = .04$, $\Delta df = 1$, $p = .84$, nor Respondent’s Sex, $\Delta\chi^2 = 1.47$, $\Delta df = 1$, $p = .10$ nor the Sex of Magician, $\Delta\chi^2 = .79$, $\Delta df = 1$, $p = .37$. Even though our participants did think that Nicolas’ tricks were better than Nathalie’s, they did not seem to be better at guessing them (NICOLAS: $M = 3.80$, $SD = 2.00$, 95% CI $[3.57, 3.95]$); NATHALIE: $M = 3.76$, $SD = 2.09$, 95% CI $[3.61, 3.98]$).

Complementary Analyses

Overall, these results support Hypothesis 1 and show that the magician’s perceived gender influences how people evaluate a magic trick: our male magician was considered better than the female one. As this effect was central to our hypothesis and as there has always been some level of controversy when examining such biases (e.g., [Swim et al., 1989](#)), we decided to calculate a Bayes factor on the effect of the perceived gender on how good the tricks were to assess the relative strength of our evidence. As such, we wanted to evaluate whether our data were sufficiently sensitive to truly support H1 (a gender effect) over H0 (no gender effect) ([Dienes, 2014, 2016](#)). In order to determine the evidence for H1 versus H0, the plausible range of effect sizes is needed. We took the women’s score as defined by how much *room to move* there was for the men’s score, in order for the latter to be evaluated as better ([Dienes, 2018a](#)). Women scored approximately 4 on a 1-7 scale; thus, the men could score between 4 and 7 by way of being rated more highly than women. That is, the maximum difference we could obtain between men and women was 3 units. [Dienes \(2018a\)](#) recommends using a Half-Cauchy distribution with *scale factor maximum* of 7 where effect sizes are expected to be relatively small. That is, following this heuristic, the scale factor would be $3/7 = 0.43$. Following the R procedure by [Dienes \(2018b website\)](#), our Bayes factor was calculated using a half-Cauchy distribution with a scale factor of .43. We used the difference of .24 as our sample mean, and SE of .10 (i.e., the raw difference divided by the t -value given by our model). Using the conventional cut-off of 3 suggested by [Jeffreys \(1961\)](#), the resulting Bayesian analysis showed evidence for the existence of a perceived gender effect over the null hypothesis, $B = 5.53$ ($B = 4.18$ for *how impressive* the trick was).

Discussion

We argue that such a gender effect is the direct consequence of the gender stereotypes associated with magicians, and the related social beliefs as to magicians’ competencies, as

predicted by Social Role Theory (Eagly, 1987). Namely, some people may believe – and have internalized – that there are more male magicians than female magicians *because* male magicians are more competent, leading them to believe that any male magician is generally better and more impressive than any female one. This effect illustrates the pernicious effects of stereotypes, leading to prejudiced evaluations.

As we gave very limited information on the magicians (i.e., just their names), it could be the case that our bias may vary, depending on the context in which it occurs. Such a variance has been shown in previous studies (e.g., Colley et al., 2003; Swim et al., 1989). As such, we wanted to examine a particularity inherent to magic (i.e., trying to spot the solution of the trick), and see whether it could lessen the gender bias we found. In a sense, we wanted to give participants some sort of opportunity to justify their lower evaluation scores. To our knowledge, this has never been done before. In reference to the present experiment, regardless of whether participants claimed they had or had not guessed the tricks, they were not asked to provide any concrete solutions to explain how the tricks were done. Thus, they were not exposed to any dissonance between their biased judgment (the female magician is less impressive) and the difficulty of providing a convincing explanation to unfamiliar tricks (especially if they did not have any). Note that even when participants claimed to have guessed the tricks in the present experiment, we cannot be sure that this was the case. Some participants may have even come up with generally superficial and easy – hence wrong – explanations (e.g., the use of a rigged deck), derived from some sort of shallow processing.

In this vein, one could argue that a way to alleviate the gender effect found in Experiment 1 would be to force participants to come up with an (in-depth) explanation after each trick. Having to provide an explanation for something for which no (easy) explanation can be found may trigger a cognitive dissonance. As such, people may find it difficult to evaluate a trick negatively when they do not know how the trick was performed (i.e., two incompatible thoughts), leading participants to judge all tricks as “better”, legitimizing the difficulty in providing a convincing solution. In the next experiment, we therefore asked participants to generate a possible solution as to how each trick was performed. We hypothesized that this would alleviate the gender bias found in Experiment 1.

Experiment 2

Method

Participants

One hundred and seventy-three Psychology undergraduate students from the University of Fribourg took part in this experiment (*Mean age* = 21.4; *SD* = 3.91; 107 women) in two different batches (N_1 = 91 from 2017; N_2 = 82 from 2018; see below for details). All stu-

dents were granted course credit for participation and were part of two different research method classes (i.e., two different years) – none had participated in the first experiment. All participants granted their written informed consent before the experiment.

Design and Procedure

The design and procedure of this experiment was the same as in Experiment 1, except that we asked a fourth question after each trick: *Even if you have not guessed the trick, please provide us with an explanation as to how the trick was performed.* Participants simply had to type in their answer.

Before the analyses, participants' explanations were coded by two independent male magicians from Besançon, following the code: 1 = *good answer / answer with the key element / possible answer*, 0 = *wrong answer / answer not precise enough*. Participants' explanations were presented randomly to the coders, who were blind to the experimental conditions. There was high agreement between the two magicians, $\kappa = .89$, $p < .001$. For the 5.32% disagreement (i.e., 124 answers out of 2331), a third magician discussed the answers with the initial coders, until all coders agreed on a final coding.

Results

Again, as in Experiment 1 and as justified by the design and convergence, we included Sex of Magician as a random slope for items. As in Experiment 1, although participants did not notice the change of magician within each session, the quality of the magic tricks was not completely homogeneous, as shown by an overall slightly better – yet non-significant – evaluation of our male magician (male magician: $M = 4.42$, $SD = 0.75$; female magician: $M = 4.19$, $SD = 1.01$, $t(13) = 1.40$, $p = .185$, 95% $CI_{\text{difference}} [-0.59, 0.13]$). Finally, again, to obtain p -values for our final model, we used the `summary()` function from the `lmerTest` package (Kuznetsova et al., 2017).

Results for the Question “How Good Was the Trick?”

Figure 1 shows the mean rating as a function of the magician's perceived gender. We compared our random model – encompassing items and participants as random intercept and Sex of the Magician as random slope for items – to one also including Perceived Gender of Magician (NATHALIE vs. NICOLAS). The latter did not show a better fit, $\Delta\chi^2 = 0.24$, $\Delta df = 1$, $p = .63$ (NATHALIE: $M = 4.27$; $SD = .943$, 95% $CI [4.08, 4.46]$; NICOLAS: $M = 4.34$; $SD = .94$, 95% $CI [4.13, 4.56]$). In fact, no other factor (i.e., Sex of Respondent or Sex of magician as fixed factors) improved the random model.

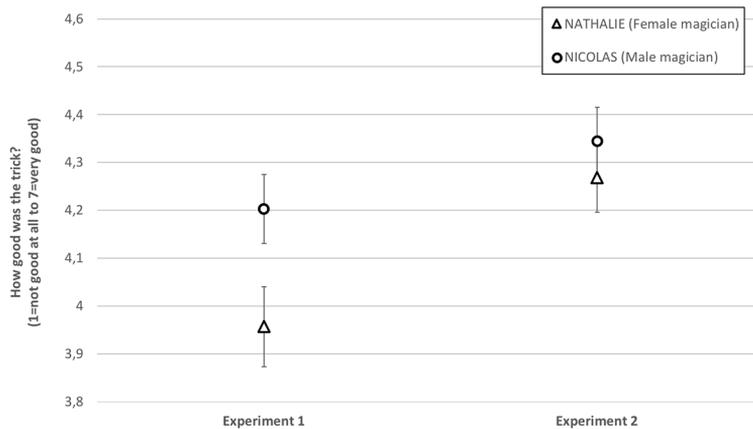


Figure 1. Comparison between Experiment 1 and Experiment 2 for the effect of Gender of Magician (i.e., Nathalie vs. Nicolas).

Note that the initial analysis was conducted on 91 participants (first batch from 2017). Since the model including Gender did not show a better fit with 91 participants, $\Delta\chi^2 = .84$, $\Delta df = 1$, $p = .36$, we decided to calculate a Bayes factor to assess the relative strength of our evidence and determine whether the non-significant effect of our Perceived Gender of Magician factor was due to data insensitivity (e.g., lack of statistical power) or was true support of the null hypothesis over the alternative hypothesis (in this case, the results of Experiment 1) (Dienes, 2014, 2016). We followed the same procedure as in Experiment 1, and for simplicity and coherence used the same scale factor of .43. As suggested by Jeffreys (1961), the Bayesian analysis showed that the data were clearly insensitive to detect either the null hypothesis or the alternative one, $B = .73$ ($B = .49$ for *how impressive* the trick was). We therefore decided to run another batch of participants (second batch from 2018) to see whether we could get enough power to detect either the null hypothesis or the alternative one, as advocated by Dienes (2014). With an additional 82 participants, our Bayesian analysis showed that the data were approaching evidence of the null hypothesis over the difference found in Experiment 1, $B = .41$ ($B = .61$ for *how impressive* the trick was), partly supporting Hypothesis 2.

Although a statistical comparison between experiments is rather delicate, as the factor *with vs. without explanation* was not run within the same experiment, we conducted an analysis on a potential interaction between experiments and the Perceived Gender of Magician. We compared a random model – encompassing items and participants as random intercept and Sex of the Magician as random slope for items – to one including Perceived Gender of Magician (NATHALIE vs. NICOLAS) and Experiment (Experiment 1 vs. Experiment 2). The latter did not show a better fit, $\Delta\chi^2 = 2.72$, $\Delta df = 3$, $p = .437$. We still computed the latter model to extract the necessary values to calculate a Bayes factor, fol-

lowing the same strategy as for the main effects of each experiment. The Bayes factor, $B = .78$, suggested that the data were neither sufficiently sensitive to substantiate H_0 (there is no interaction between experiments) nor H_1 (there is an interaction between experiment). We will come back to this issue in the Discussion section.

Results for the Question “How Impressive Was the Trick?”

The correlation between the first and second measure (*How good is the trick?* and *How impressive is it?*) was again very high, $r(171) = .89$, 95% CI [.854, .917], $p < .001$. Not surprisingly then, when comparing our random model – encompassing items and participants as random intercept and Sex of the Magician as random slope for items – to one also including Perceived Gender of Magician (NATHALIE vs. NICOLAS), the latter did not show a better fit, $\Delta\chi^2 = 0.844$, $\Delta df = 1$, $p = .358$ (NATHALIE: $M = 4.00$; $SD = .963$, 95% CI [3.805, 4.196]; NICOLAS: $M = 4.14$; $SD = .984$, 95% CI [3.918, 4.365]) ($B = .61$). As in Experiment 1, no other factor (i.e., Sex of Respondent or Sex of magician as fixed factors) improved the random model.

Results for the Questions “Did You Guess the Magic Trick? And “How Was It Done?”

The correlation between the first and third question (*How good is the trick?* and *Did you guess how the trick was performed?*) was low and negative, yet significant, $r(170) = -.369$, 95% CI [-.491, -.232], $p < .001$. As in Experiment 1, we present the same analyses as for the first two questions. In addition to the analysis of the question *Did you guess the magic trick?*, we also analyzed participants’ actual response accuracy.

Although adding Perceived Gender of Magician did not improve the initial random model, Sex of Respondent did, $\Delta\chi^2 = 8.422$, $\Delta df = 1$, $p = .004$. The model estimates are shown in Table 3. These results showed that women were less likely to claim to have correctly guessed the secret, $M = 3.44$, $SD = .971$, 95% CI [3.26, 3.63], than men, $M = 3.89$, $SD = .983$, 95% CI [3.650, 4.13], $t(170) = -2.931$, $p = .004$. Adding any other factor did not further improve the model.

Table 3

Model Estimates for Our Best Fitting Model for How Participants Claim to Have Guessed the Tricks

Factor	Estimate	SE	df	t	Pr(> t)
(Intercept)	3.901	0.228	23.17	17.09	< .001
Sex of respondent (female)	-0.449	0.153	170.32	-2.93	.004

Note. Includes Sex of Respondent as a fixed factor, participants and items as random intercepts, and Sex of Magician as random slope for items. Treatment contrasts were used. The model including Gender of Magician was significant in Experiment 1, whereas it was not in Experiment 2.

Of interest, there was a significant correlation between participants claiming to have guessed how the trick was done (i.e., *Did you guess how the trick was performed?*) and their actual knowledge of the secret, $r(170) = .320$, $p < .001$, 95% CI [.179, .448]. Male participants also generated more correct explanations than female participants (Female participants: $M = .364$, $SD = .135$, 95% CI [.338, .390]; Male participants: $M = .409$, $SD = .142$, 95% CI [.374, .444], $t(132) = 2.02$, $p = .045$). We will come back to this effect in the General discussion section.

General Discussion

In two experiments, we presented participants with magic tricks and asked them to evaluate each of them in terms of how good and how impressive they were. In Experiment 2, participants were additionally required to generate possible solutions as to how each of the tricks was done. In both experiments, following the procedure introduced by Goldberg's (1968) seminal work, half of the participants were made to believe that the tricks were performed by a woman, and half by a man. This manipulation was independent of the magician's true gender. In Experiment 1, participants felt that the tricks performed by a man were better than those allegedly performed by a woman, which supported our predicted gender effect (Hypothesis 1). This is in line with Goldberg's (1968) initial results, showing that participants were more critical of journal articles written by women than by men, especially when the content of the articles was related to male-dominated fields. However, once participants had to generate possible explanations about how the tricks were done (Experiment 2), the data suggested that the gender difference found in Experiment 1 may be fluctuant, and depend on different evaluation strategies, partly supporting Hypothesis 2.

Social Role Theory

The results of Experiment 1 are in line with Social Role Theory. Our participants – in line with the history of magic – may have internalized the belief that men have better dispositions for magic than women, and as such are better performers, irrelevant of their *actual* performance. We could further claim that our participants in Experiment 1 faced some level of incongruence – following Eagly and Karau's (2002) Role Congruity Theory – between a *masterful, competent* magician and a woman. Blending both information may well be challenging, hence the female magician was judged as less competent. Note that even if the female magician had been evaluated as similarly competent to the male one, Role Congruity Theory would predict that female magicians are still evaluated less favorably on other dimensions (e.g., more emotional). This could actually be the case in Experiment 2, although we have no data (yet) to warrant such an idea.

Still, the different patterns of results between Experiment 1 and Experiment 2 do hint at the possibility that different mental mechanisms, or strategies were at stake in these experiments. Applying different strategies when making judgments is reminiscent of the idea that we often make judgments based on simplifying strategies (i.e., heuristics), rather than on extensive algorithmic processing (e.g., Parzuchowski, Bocian, & Gygax, 2016; Kahneman & Tversky, 1996). Indeed, as presented in the introduction, there are numerous other domains in which gender stereotypes influence people's judgments. However, in the current paper we begin to show that when pushed towards making more analytical judgments (as in Experiment 2), superficial and simplified information (i.e., stereotypes) might be less influential.

The Role of Cognitive Dissonance

Although a more detailed description of the mechanisms involved is only speculative at this point – the data in Experiment 2 were not entirely conclusive –, we still would like to suggest a possible explanation. Since the reduction in gender bias was based on mechanisms not directly related to the actual magician, our explanation deviates from those that focus on individuation (i.e., giving more information about the protagonists than just the name, as in Colley et al., 2003). As such, we propose that the seemingly attenuated gender effect reported in Experiment 2 may be rooted in some sort of reduced dissonance. Nardi (1988) has argued that people often feel powerless when tricked by conjuring effects, and that they feel cognitively challenged by the magician. Being fooled and impressed by a magician may therefore result in a cognitive conflict that is decreased if we presume that the trick was performed by a good magician. As such, participants may be more comfortable to admit to having been fooled by someone stereotyped as a *good* magician (the male magician) than by someone stereotyped as an *inferior* one (the female magician). Consequently, in Experiment 1, participants evaluated the female magician as less impressive than the male one. In Experiment 2, participants were under pressure, as they were confronted with two opposing drives: on the one hand, being confronted with a female (inferior) magician, yet on the other hand not being able to understand how they were fooled, whilst actively searching for a solution. The latter most likely overpowered the former to reduce the dissonance, resulting in better evaluations – at least numerically – of the female magician's tricks in Experiment 2 than in Experiment 1. One could also argue that forcing participants to come up with possible solutions to tricks that they just evaluated may push them to be *accountable* for their evaluations. As such, when participants give any score, they become accountable for giving it, and only by giving a correct solution can they warrant a negative evaluation. Based on the seminal work by Tetlock and colleagues on accountability (e.g., Tetlock, Skitka, & Boettger, 1989), some studies have actually shown that gender biases disappear when participants are made accountable for their decisions in professional settings (e.g., Girvan, Deason, & Borgida, 2015).

However, more research is needed to provide more substantial evidence of this idea. Future research might focus on possible dissonance mechanisms by measuring the relative discomfort generated by being fooled by a female magician without understanding how the trick was done and being asked to explain the trick. Also, our data cannot dismiss the idea that the gender bias found in Experiment 1 was only apparent because reading the magician's name was the only *other* task besides simply watching the trick. In this sense, it could be the case that a more resource-demanding task (i.e., trying to find a solution) may exhaust all available resources, preventing participants from actually considering gender as a relevant cue. This could well explain the fact that many studies (e.g., Colley et al., 2003) did not find gender bias when additional and individuating information regarding the target protagonist was given.

The Role of Confidence

As a final note, we found it interesting that male participants were more confident that they had correctly guessed how the tricks were done, and they also came up with more correct solutions. These results dovetail Nardi's (1988) observation that "Although most people respond similarly to a trick, men more often than women state they also know one, or publicly attempt to figure out how it was done". One tentative explanation of such a gender effect is that it may arise from the gender stereotype *boosting* men's confidence to find the right solution but *lowering* confidence in women. The activation of the gender stereotype associated with magic may well activate lower self-efficacy in women, through the well-documented mechanism of stereotype threat, mimicking effects found in computing skills research (e.g., Christoph, Goldhammer, Zylka, & Hartig, 2015) or mathematics (e.g., in Spencer, Steele, & Quinn, 1999). Consequently, and following the expectancy-value theory of achievement motivation (e.g., Eccles & Wigfield, 2002), women may not feel as legitimate as men to generate possible solutions. In fact, this may also explain the rather low number of women in magic, especially if women – as in other domains that are stereotypically male (Dickhäuser & Stiensmeier-Pelster, 2002) – attribute their failures more to internal, global, stable and uncontrollable factors than they do for their successes.

Conclusion

To sum up, there are many factors that can contribute to the appreciation of magic, and we have shown that some of these factors could be grounded in social biases, such as gender stereotypes. However, neither of our experiments allows us to determine exactly at which temporal level of magic trick appreciation the suggested effects takes place. Namely, to evaluate a magic trick, different cognitive (and social) processes are at stake, and it is yet difficult to determine which one is most sensitive to social biases.

To conclude, whatever process is at stake here, female magicians have it harder (and are prejudiced to refer back to Goldberg's discussion), and identical performances may be less appreciated. Our findings dovetail Bruns and Zompetti's (2014) observation: "Women in magic are still "muted" in the mainstream magic entertainment circuit and seen as a novelty". However, our study suggests a possible way in which these gender stereotypes can be alleviated, although the data are not yet entirely conclusive. We still hope that future research will concentrate on all possible ways to alleviate gender biases during magical performances, and the world more generally.

Funding: The authors have no funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: We would like to express our sincere gratitude to Zoltan Dienes for providing us with very clear guidance regarding Bayes factors.

Ethics Approval: The experiments reported here were approved by the Ethical Committee of the Department of Fribourg, Switzerland, and participants gave their written consent.

Data Availability: Datasets for the studies are freely available (see the [Supplementary Materials](#) section).

Supplementary Materials

The following Supplementary Materials are available (for access see [Index of Supplementary Materials](#) below):

- Via the OSF repository: The original data and analysis R scripts of Experiments 1 and 2
- Via the PsychArchives repository: All videos (*Note:* Although some videos have sound, they were muted in the experiments)

Index of Supplementary Materials

Gygax, P., Thomas, C., Didierjean, A., & Kuhn, G. (2019). *Supplementary materials to "Are women perceived as worse magicians than men? Gender bias when evaluating magic tricks"* [Data and R scripts]. OSF. <https://osf.io/tqxs5>

Gygax, P., Thomas, C., Didierjean, A., & Kuhn, G. (2019). *Supplementary materials to "Are women perceived as worse magicians than men? Gender bias when evaluating magic tricks"* [Videos]. PsychOpen. <https://doi.org/10.23668/psycharchives.2629>

References

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, United Kingdom: Cambridge University Press.
- Bruns, L. C., & Zompetti, J. P. (2014). The rhetorical goddess: A feminist perspective on women in magic. *Journal of Performance Magic*, 2(1), 8-39. <https://doi.org/10.5920/jpm.2014.218>
- Brybaert, M. (2007). "The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem (Report written for RTN-LAB, Version 2.0). Egham, United Kingdom: Royal Holloway, University of London.
- Chollet, M. (2018). *Sorcières – La puissance invaincue des femmes* [Witches – The undefeated power of women]. Paris, France: La Découverte.
- Christoph, G., Goldhammer, F., Zylka, J., & Hartig, J. (2015). Adolescents' computer performance: The role of self-concept and motivational aspects. *Computers & Education*, 81, 1-12. <https://doi.org/10.1016/j.compedu.2014.09.004>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335-359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3)
- Colley, A., North, A., & Hargreaves, D. J. (2003). Gender bias in the evaluation of New Age music. *Scandinavian Journal of Psychology*, 44, 125-131. <https://doi.org/10.1111/1467-9450.00330>
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29(1), 80-85. <https://doi.org/10.1037/h0035733>
- Dickhäuser, O., & Stiensmeier-Pelster, J. (2002). Gender differences in computer work: Evidence for the model of achievement-related choices. *Contemporary Educational Psychology*, 27(3), 486-496. <https://doi.org/10.1006/ceps.2001.1106>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, Article 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78-89. <https://doi.org/10.1016/j.jmp.2015.10.003>
- Dienes, Z. (2018a). *How do I know what my theory predicts?* Talk given at University of Sussex on 5 Dec 2018 [Video file]. Retrieved from https://www.youtube.com/watch?v=_NRaNKhtVOK
- Dienes, Z. (2018b). *Making the most of your data with Bayes*. Retrieved from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm
- Eagly, A. H. (1987). *John M. MacEachran memorial lecture series; 1985. Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.
- Eagly, A. H. (1997). Sex differences in social behavior: Comparing social role theory and evolutionary psychology. *American Psychologist*, 52(12), 1380-1383. <https://doi.org/10.1037/0003-066X.52.12.1380.b>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573-598. <https://doi.org/10.1037/0033-295X.109.3.573>

- Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, 17(3), 306-315.
<https://doi.org/10.1177/0146167291173011>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA, USA: Stanford University Press.
- Field, A. (2014). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, United Kingdom: Sage.
- Fleischmann, A., Sieverding, M., Hespeneide, U., Weiß, M., & Koch, S. C. (2016). See feminine – Think incompetent? The effects of a feminine outfit on the evaluation of women’s computer competence. *Computers & Education*, 95, 63-74. <https://doi.org/10.1016/j.compedu.2015.12.007>
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, 40, 206-212. <https://doi.org/10.3758/BRM.40.1.206>
- Girvan, E. J., Deason, G., & Borgida, E. (2015). The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions. *Law and Human Behavior*, 39, 525-537. <https://doi.org/10.1037/lhb0000139>
- Goldberg, P. (1968). Are women prejudiced against women? *Trans-Action*, 5(5), 28-30.
<https://doi.org/10.1007/BF03180445>
- Green, L. (1997). *Music gender, education*. Cambridge, United Kingdom: Cambridge University Press.
- Jeffreys, H. (1961). *Theory of probability*. Oxford, United Kingdom: Oxford University Press.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103, 582-591. <https://doi.org/10.1037/0033-295X.103.3.582>
- Koch, A. J., D’Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100, 128-161. <https://doi.org/10.1037/a0036734>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1-26.
<https://doi.org/10.18637/jss.v082.i13>
- Mangan, M. (2007). *Performing dark arts*. Chicago, IL, USA: Intellect Books.
- Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., . . . Szesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, 46(2), 841-871.
<https://doi.org/10.3758/s13428-013-0409-z>
- Nardi, P. M. (1988). The social world of magicians: Gender and conjuring. *Sex Roles*, 19(11-12), 759-770. <https://doi.org/10.1007/BF00288991>
- Öttl, A., & Behne, D. M. (2016). Experience-based probabilities modulate expectations in a gender-coded artificial language. *Frontiers in Psychology*, 7, Article 1250.
<https://doi.org/10.3389/fpsyg.2016.01250>

- Parzuchowski, M., Bocian, K., & Gygax, P. (2016). Sizing up objects: The effect of diminutive forms on positive mood, value, and size judgments. *Frontiers in Psychology, 7*, Article 1452. <https://doi.org/10.3389/fpsyg.2016.01452>
- Proudfoot, D., Kay, A. C., & Koval, C. Z. (2015). A gender bias in the attribution of creativity: Archival and experimental evidence for the perceived association between masculinity and creative thinking. *Psychological Science, 26*, 1751-1761. <https://doi.org/10.1177/0956797615598739>
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria.
- Sczesny, S., Spreemann, S., & Stahlberg, D. (2006). Masculine = competent? Physical appearance and sex as sources of gender-stereotypic attributions. *Swiss Journal of Psychology, 65*, 15-23. <https://doi.org/10.1024/1421-0185.65.1.15>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28. <https://doi.org/10.1006/jesp.1998.1373>
- Swim, J. K., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*(3), 409-429. <https://doi.org/10.1037/0033-2909.105.3.409>
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability – Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology, 57*, 632-640. <https://doi.org/10.1037/0022-3514.57.4.632>



Social Psychological Bulletin is an official journal of the Polish Social Psychological Society (PSPS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology Information (ZPID), Germany.