Journal of Experimental Social Psychology 95 (2021) 104144

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

Moral credentials and the 2020 democratic presidential primary: No evidence that endorsing female candidates licenses people to favor men☆,☆

Laura M. Giurge *, Eva Hsin-Lian Lin, Daniel A. Effron

London Business School, United Kingdom

ARTICLE INFO

Editor: Rachel Barkan

Keywords:
Moral credentials
Moral licensing
Gender bias
Sexism
Voting

ABSTRACT

Endorsing Obama in 2008 licensed some Americans to favor Whites over Blacks—an example of *moral self-licensing* (Effron, Cameron, & Monin, 2009). Could endorsing a female presidential candidate in 2020–21 similarly license Americans to favor men at the expense of women? Two high-powered, pre-registered experiments found no evidence for this possibility. We manipulated whether Democrat participants had an opportunity to endorse a female Democratic candidate if she ran against a male candidate (i.e., Trump in Study 1, $N = 2143$; an anti-Trump Republican or independent candidate in Study 2, $N = 2228$). Then, participants read about a stereotypically masculine job and indicated whether they thought a man should fill it. Contrary to predictions, we found that endorsing a female Democrat did not increase participants' tendency to favor men over women for the job. We discuss implications for the robustness and generalizability of moral self-licensing.

In 2020, more women than ever before campaigned to be the presidential nominee of a major American political party. Some commentators pointed to the three female candidates as a sign of progress towards gender equality. Others—highlighting Hillary Clinton's historic loss to Donald Trump four years earlier—worried about gender bias. Indeed, bias towards members of underrepresented groups often accompanies progress towards equality (e.g., Georgeac & Rattan, 2019; Kaiser et al., 2013). For example, during the 2008 U.S. presidential election, Obama supporters who were given a chance to express their support for Obama became more likely to subsequently endorse ambiguous views that favored Whites at the expense of Blacks (Effron et al., 2009). Supporting a Black presidential candidate made people feel they had earned *moral credentials* as unprejudiced, enabling them to favor Whites without worrying about seeming racist (Monin & Miller, 2001). The present research investigates whether gender diversity in the 2020 Democratic primaries could have a similar ironic effect. Could endorsing a female presidential candidate license voters to subsequently favor men at the expense of women?

People prefer to feel and appear non-prejudiced (Crandall & Eshleman, 2003). Establishing evidence that they are non-prejudiced frees people to express views that would otherwise cast aspersions on their egalitarianism (Bradley-Geist, King, Skorinko, Hebl, & McKenna, 2010; Cascio & Plant, 2015; Effron, 2014; Effron, Miller, & Monin, 2012; Kouchaki, 2011; Mann & Kawakami, 2012; Merritt et al., 2012; Monin & Miller, 2001; Simon & O'Brien, 2015). This phenomenon exemplifies *moral self-licensing,* whereby doing "good" can disinhibit people to do "bad" (for reviews, see Effron, 2016; Effron & Conway, 2015; Klotz & Bolino, 2013; Merritt, Effron, & Monin, 2010; Miller & Effron, 2010; Mullen & Monin, 2016). Prior work argues that imagining or performing virtuous behaviors can license less prosocial behavior (Conway & Peetz, 2012; Jordan, Mullen, & Murnighan, 2011), more cheating (Clot, Grolleau, & Ibanez, 2014; Mazar & Zhong, 2010), more indulgent consumption choices (Effron, Monin, & Miller, 2013; Khan & Dhar, 2006; Schwabe, Dose, & Walsh, 2018), less environmentally friendly behaviors (Gholamzadehmir, Sparks, & Farsides, 2019; Meijers, Verlegh, Noordewier, & Smit, 2015; Tiefenbeck, Staake, Roth, & Sachs, 2013), and more workplace deviance (List & Momeni, 2020; Loi, Kuhn, Sahaym, Butterfield, & Tripp, 2020; Yam, Klotz, & Reynolds, 2017). Overall, doing good allows people to feel they have proven themselves to be adequately moral, enabling them to give into temptations without feeling or appearing too unethical. Similarly, non-prejudiced acts make people feel they have proven themselves to be non-prejudiced (Effron, 2014), disinhibiting subsequent prejudiced actions (Miller & Effron, 2010).

☆ Data, code, experimental materials, and pre-registration documents are available via the Open Science Framework: https://osf.io/ja83x/.☆ This paper has been recommended for acceptance by Professor. Rachel Barkan.

* Corresponding author at: London Business School, R309, Regent's Park, London NW1 4SA, United Kingdom.
E-mail address: lgiurge@london.edu (L.M. Giurge).

https://doi.org/10.1016/j.jesp.2021.104144
Received 21 June 2020; Received in revised form 23 March 2021; Accepted 31 March 2021
Available online 23 April 2021

The 2020 U.S. Democratic presidential primary provides a unique opportunity to test the robustness and generalizability of moral self-licensing in a consequential, contemporary context. As noted, prior work found evidence that endorsing Obama licensed his supporters to favor Whites over Blacks (Effron et al., 2009). Yet, to our knowledge, no studies have assessed whether a similar licensing effect would emerge when candidate gender, rather than race, is salient. A classic critique of social psychology is that its theories rarely generalize beyond the historical context where they are developed (Gergen, 1973). Is the licensing effect that was observed with Obama specific to a particular moment in America's political history, or does it emerge when members of a different underrepresented group run for public office?

Endorsing a woman in the 2020 Democratic primaries could license voters to subsequently favor men at the expense of women. In prior work, rejecting sexist statements licensed people to subsequently say that a stereotypically masculine job is better suited for men than women (Monin & Miller, 2001) – an effect replicated in a high-powered, pre-registered, multi-lab study (Ebersole et al., 2016). In 2020, supporting a woman for one of the most powerful jobs in the world—U.S. President—might also feel like a "non-sexist credential" that licenses people to subsequently favor men over women. Expressing an intention to vote for a woman over President Donald Trump – the Republican nominee in 2020 – could feel like a particularly strong non-sexist credential to Democrats, given allegations of sexism against Trump (e.g., Chozick & Parker, 2016; Filipovic, 2017) and perceptions that sexism tilted the 2016 election in his favor (see Glick, 2019; Ratliff, Redford, Conway, & Smith, 2019; Schaffner, MacWilliams, & Nteta, 2018).

However, two key limitations of the moral licensing literature cast doubt on whether a licensing effect would emerge in contemporary American politics. First, although moral licensing has been found in diverse situations, prior theorizing has not sufficiently specified what behaviors "count" as a license in people's minds (see Effron, 2016; Effron & Conway, 2015; Mullen & Monin, 2016). Thus, it is unclear to which novel situations these results generalize. Expressing an intention to vote for a female president in 2020 may *not* feel like a non-sexist credential in the same way that voting for a Black president in 2008 felt like a non-racist credential.

Second, not all documented examples of moral licensing have proven robust. Although as noted the gender-licensing effect originally observed by Monin and Miller (2001) has replicated (Ebersole et al., 2016), findings from other paradigms unrelated to discrimination or sexism have not (Blanken, van de Ven, Zeelenberg, & Meijers, 2014; Jordan et al., 2011; Mazar & Zhong, 2010; Rotella & Barclay, 2020; Sachdeva, Iliev, & Medin, 2009; Urban, Bahník, & Kohlová, 2019). Underpowered studies and publication bias have both contributed to this issue (Blanken et al., 2014; Kuper & Bott, 2019; Simbrunner & Schlegelmilch, 2017).

Thus, to support the prior claim that moral licensing is a common, generalizable effect, it would be necessary to find robust evidence of licensing across high-powered experiments in multiple and real-world contexts. Although prior work found evidence of licensing in the context of Obama's election (Effron et al., 2009), subsequent work suggests that the sample size in these studies would be too small to detect a typically sized moral licensing effect (e.g., Effron et al., 2009, Study 1's $N = 84$ across two cells would have only 41% power to detect, with a one-tailed test, the average effect size of $d = 0.31$ documented in the licensing literature; Blanken, Van de Ven, & Zeelenberg, 2015). No other studies that we know of have examined licensing in politics in over a decade.

The present research provides two high-powered tests of licensing during and shortly after the 2020 Democratic primaries. Millions of American voters cast their ballots for a female presidential candidate during those primaries, and tens of millions more did so in the 2016 Presidential election. If our experiments found that endorsing a woman for president could license people to subsequently favor men over women, it would not only provide evidence for moral licensing's robustness and generalizability; it would also point to a troubling, ironic consequence of gender diversity in politics. Thus, our main goal was to test the following hypothesis:

H1. Giving Democrats a chance to endorse a female presidential candidate will increase their willingness to express a view that favors men at the expense of women.

A secondary goal was to clarify whether certain individual differences moderate moral licensing. In theory, people with more prejudiced attitudes should be more inclined to use a non-prejudiced behavior as a license to express potentially problematic views, whereas people with less prejudiced attitudes should not be tempted to express such views even after acting in a non-prejudiced way. Indeed, among Obama supporters, those with greater racial prejudice were more likely to use their endorsement of Obama as a license (Effron et al., 2009). However, the effect of other manipulations of non-racist credentials did not significantly depend on racial attitudes (Effron et al., 2012). To better understand the relationship between individuals' prejudice and licensing in a gendered context, we included multiple measures of sexism (Glick & Fiske, 1996; Swim, Aikin, Hall, & Hunter, 1995) and tested the following hypothesis:

H2. The effect described in H1 will be stronger among Democrats with more-sexist attitudes and weaker among those with less-sexist attitudes.

We also explored two other individual differences that could moderate licensing in the present context. First, we examined participants' motivation to respond without prejudice (Klonis, Plant, & Devine, 2005). Prior work has examined this variable, but the sample sizes were too small to draw strong conclusions (Effron et al., 2012; Monin & Miller, 2001). Second, we considered participant gender. In theory, men could be more likely than women to use past behavior as a license to subsequently express sexist views (see Monin & Miller, 2001). However, prior work has not consistently found support for this possibility (Ebersole et al., 2016).

We tested our hypotheses in two pre-registered, large-sample experiments (combined $N = 4371$) – among the largest moral licensing studies ever conducted (see Blanken et al., 2015; Simbrunner & Schlegelmilch, 2017). In Study 1, conducted during the 2020 U.S. Democratic primaries, Democrats were randomly assigned an opportunity to endorse a female Democratic candidate over Donald Trump, an opportunity to endorse a male Democratic candidate over Trump, or no such opportunity. Based on the moral self-licensing literature, we expected that endorsing a female candidate would make participants feel like they had established non-sexist credentials, thus licensing them to favor men over women on a subsequent hiring task. The results showed no evidence of this prediction. We ran Study 2 with a new sample of Democrats soon after the 2020 Presidential election to address the possibility that endorsing a woman over Donald Trump in particular – or over a male Republican in general – had somehow prevented a licensing effect from emerging in Study 1. The results showed no evidence of that possibility. Specifically, endorsing a woman over a man for a potential 2024 presidential bid had no measurable licensing effect – regardless of whether the man was an anti-Trump Republican or a political independent candidate. Both experiments had enough statistical power to detect even a small effect.

We report all measures, conditions, exclusions, and the method for determining the final sample size. Verbatim materials, data, code, and the pre-registrations for both studies are available at: https://osf.io/ja83x/.

1. Study 1

1.1. Method

We pre-registered this study at aspredicted.org/blind.php?x=a3q2kr and collected data between March 3rd–7th, 2020.

1.1.1. Participants

We powered Study 1 to detect the effect size observed in a high-powered, pre-registered study that used the same dependent measure as we did: $d = 0.14$ (Ebersole et al., 2016).[1] Detecting this effect size between two conditions 85% of the time and using a one-tailed test requires 735 people per cell (Faul, Erdfelder, Lang, & Buchner, 2007); our design had three cells, so, anticipating data exclusions, we targeted 2400 participants (i.e., 800 per cell).

We recruited this sample on Prolific Academic, a higher-quality online participant panel than Amazon Mechanical Turk (Peer, Brandimarte, Samat, & Acquisti, 2017). Using Prolific's pre-screen filters, we targeted American Democrats who voted for Hillary Clinton in 2016. Participants were not informed of these pre-screen criteria. To improve data quality, we used filters in Qualtrics to prevent participation by people who failed a reading-comprehension question, accessed the survey on a mobile device, or resided outside the U.S.

Of the 2459 people who started the study, 2376 provided full responses, and we further excluded 257 people according to our pre-registered exclusion criteria: an IP address outside the U.S., a duplicate IP address, a duplicate Prolific ID, or saying at the end of the study that they were not a Democrat. Our final dataset contained 2143 participants (60.40% women; $M_{\text{age}} = 35.83$ years, $SD = 12.47$; see Supplementary Material, Table S1 for means, standard deviations, and correlations). At the end of the study, almost all participants confirmed their pre-screen responses that they had voted in the 2016 U.S. presidential election (94.60%) and they had or would have voted for Hillary Clinton (98.70%). A sensitivity analysis confirmed that the final sample size allowed us to detect a moral licensing effect of $d = 0.14$ with 85% power ($\alpha = 0.05$ by a pre-registered one-tailed test).

1.1.2. Procedure

We randomly assigned participants to one of three conditions (adapted from Effron et al., 2009). In the endorse-woman ($n = 716$) and endorse-man ($n = 713$) conditions, participants viewed a Democrat's name and picture next to Donald Trump's name and picture (display order randomized) and indicated whom they would vote for if those two candidates ran against each other for president. (We listed both candidates' political party.)

In the endorse-woman condition, we randomized the Democrat to be either Elizabeth Warren or Amy Klobuchar—the top two female candidates at the time of the study (Warren suspended her campaign on March 5, 2020). In the endorse-man condition, we randomized the Democrat to be either Bernie Sanders or Tom Steyer because their standing in the polls was most similar to Warren and Klobuchar. In the control condition ($n = 714$), participants responded to the dependent measure (described below) without endorsing any candidate.

1.2. Measures

1.2.1. Dependent measure

Participants first completed three filler questions (e.g., "Are you at all familiar with the building industry?") and then the dependent measure asking them to indicate whether they thought a stereotypically masculine job was better suited for a particular gender (from Monin & Miller, 2001; p. 35) on a 7-point scale ranging from −3 (*Yes, much better for women)* to 3 (*Yes, much better for men*), with a midpoint of 0 (*No, I do not feel this way at all*).

> "Imagine that you are the manager of a small (45-person) cement manufacturing company based in New Jersey. Last year was a particularly good one, and after you invested in increasing the output capacity of your plant, you decide that it would be very fruitful if you could find clients in other states to increase your business. Because you cannot spend too much time away from the plant, you decide to appoint someone to go around to prospective clients and negotiate contracts. This is a highly specialized market, and the job will mostly consist in going from one building site to another, establishing contacts with foremen and building contractors. It is also a highly competitive market, so bargaining may at some points be harsh. Finally, it's a very technical market, and a representative that did not exude confidence in their technical skills would not be taken seriously by potential clients. Realizing how useful such help would be for you, you decide to give the person chosen one of the top-five salaries in your company. Do you feel that this job is better suited for one gender rather than the other?"

This measure, along with an analogous measure about race, is a standard instrument in the prejudice-licensing literature (Bradley-Geist et al., 2010; Cascio & Plant, 2015; Ebersole et al., 2016; Effron et al., 2009, 2012; Monin & Miller, 2001), designed to pull participants in two directions simultaneously. Concern that a woman would underperform or experience discrimination in a male-dominated industry could lead people to prefer to hire a man rather than a woman. By contrast, expressing such a preference could make people feel or appear sexist. Expressing support for a female presidential candidate earlier in the study should resolve this tension. We predicted that participants who had just "proven" their lack of sexism with their choice of a female candidate for president would feel more comfortable expressing a stereotypical hiring preference. We emphasize that this measure was not designed to assess private gender attitudes, but rather assess willingness to publicly express a view that triggers worry about feeling or appearing sexist (Monin & Miller, 2001).

1.2.2. Potential moderators

We measured participants' gender attitudes with the 8-item modern sexism scale ($\alpha = 0.85$; e.g., "Women often miss out on good jobs due to sexual discrimination"; Swim et al., 1995), and the benevolent and hostile sexism scales[2] (11 items each; αs $= 0.88$ and 0.83, respectively; e.g., "Women should be cherished and protected by men" [benevolent], and "Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for 'equality'" [hostile]; Glick & Fiske, 1996). As an exploratory step, we also measured participants' internal and external motivation to control sexist responding (αs $= 0.92$ and 0.81, respectively; e.g., "According to my personal values, using stereotypes about women is OK" [reverse-scored internal item] and "I attempt to appear non-sexist toward women in order to avoid disapproval from others" [external item]; Klonis et al., 2005). We administered these scales after the dependent measure to avoid influencing people's responses to them (Effron et al., 2009).

1.2.3. Political preferences

To validate participants' pre-screening responses, we next asked whether participants had voted in the 2016 U.S. presidential election. Those who said 'yes' indicated for whom they had voted, whereas those who said 'no' indicated for whom they *would* have voted. Then, participants indicated the political party they identified with or leaned towards. As noted, these measures came *after* the dependent measure and moderators.

Finally, we paired Donald Trump with each of the four Democrats (i.

[1] As noted, the average effect size in the licensing literature is estimated to be larger, $d = 0.31$, but this estimate comes from a wide variety of paradigms and is likely inflated by publication bias (Blanken et al., 2015), so we powered the study to detect a smaller effect.

[2] Note that in our pre-registration, H2 focused specifically on modern sexism because prior research on licensing found moderation by a measure of modern racism (Effron et al., 2009, Study 3). Thus, in the absence of prior data on hostile and benevolent sexism, we considered these measures as exploratory.

e., Warren, Klobuchar, Sanders, and Steyer), and asked participants to indicate whom they would vote for in each pair. (In the endorse-man and endorse-woman conditions, we omitted the pair that we had asked participants about earlier in the study.)

1.3. Results and discussion

1.3.1. Manipulation check

As expected, virtually all participants endorsed the Democrat in the endorse-man condition (98.30%) and the endorse-woman condition (98.50%).

1.3.2. Outliers

As pre-registered, we replaced the 62 observations on the dependent measure that fell at or more than 3.29 *SDs* away from the grand mean ($M = 0.30$, $SD = 0.77$) with the next-smallest value ($z = 3.29$ corresponds to $p < .001$; Tabachnick, Fidell, & Ullman, 2007).

1.3.3. No support for the moral licensing hypothesis

The results did not support the moral licensing hypothesis (see Fig. 1). Our pre-registered prediction was that giving Democrats an opportunity to endorse a female presidential candidate—compared to giving them an opportunity to endorse a male candidate or giving them no such opportunity—would increase their willingness to say they would hire a man instead of a woman for the hypothetical job. That is, participants in the endorse-woman condition, compared to the endorse-man or the control condition, should express a stronger preference for hiring a man.

To test this prediction, we followed our pre-registered plan and first created two dummy variables (i.e., endorse-man and control condition), with the endorse-woman condition as the reference group. We then regressed participants' choice in the hiring scenario on both dummy variables. The results showed no evidence that participants were more likely to favor men for the job in the endorse-woman condition ($M = 0.26$, $SD = 0.63$) than in either the control condition ($M = 0.34$, $SD = 0.70$), $b = 0.08$, $SE = 0.04$, $t = 2.31$, $p = .989$[3], $d = -0.12$, 95% $CI_d = [-0.22; -0.02]$ or the endorse-man condition ($M = 0.25$, $SD = 0.68$), $b = -0.01$, $SE = 0.04$, $t = -0.17$, $p = .566$, $d = 0.02$, 95% $CI_d = [-0.09; 0.12]$, for pre-registered one-tailed tests. (The pre-registered moral licensing hypothesis in this study predicts significant negative effects for both dummy codes; Cohen's d is coded so that positive numbers indicate a licensing effect.) We emphasize that the estimated effect size in the former test was in the opposite direction than predicted, and the one in the latter test was virtually zero (i.e., $d = 0.02$, which an experiment would require 61,828 participants to detect 80% of the time with a one-tailed test).

Fig. 2 plots standardized effect sizes and their 95% *CI*s. We can be confident that endorsing a woman (vs. no such opportunity; i.e., the control condition) did not license people, because the 95% *CI* for this comparison excluded positive numbers—and if endorsing a woman (vs. a man) licensed people, the effect size would be very small (i.e., $d < 0.12$, the upper-bound of the 95% *CI*). Comparing the endorse-woman condition to the average of the other two conditions further highlights the lack of evidence for licensing. The best estimate of the size of this comparison is $d = -0.06$ (i.e., in the opposite direction as the licensing hypothesis), and the largest plausible licensing effect consistent with this estimate is $d = 0.03$ (i.e., the top of the 95% *CI*).

1.3.4. Exploratory analyses of condition differences

Exploratory analyses with two-tailed tests did find that the means for the three conditions differed significantly, $F(2, 2140) = 3.81$, $p = 0.022$ (see Fig. 1). Pairwise comparisons showed that participants were more likely to favor men for the job in the control condition, compared to the endorse-man or the endorse-woman condition, *ts* (2140) = 2.47 and 2.31, *ps* = 0.014 and 0.021 (two-tailed), *ds* = 0.13 and 0.12, respectively. However, we urge caution in interpreting these effects because (a) they were in the opposite direction of our pre-registered one-tailed tests, and (b) only the first effect remains significant after applying a Bonferroni correction for multiple comparisons (critical $\alpha = 0.016$ after the correction).

1.3.5. Bayesian analysis

The analyses thus far fail to support the licensing hypothesis. To further understand the data, we ran Bayesian regression analysis using *Stan* (Carpenter et al., 2017) and the *bridgesampling* package (Gronau, Singmann, & Wagenmakers, 2017) in R. This exploratory analysis used the same dummy codes as the pre-registered regression described above. Results suggest that the data are over three times more likely under the null model than the model with the two dummy predictors ($BF_{01} = 3.22$). Thus, the Bayesian analysis suggests that the data are more consistent with a null effect than with a licensing effect.

1.3.6. No moderation by gender attitudes, participant gender, or the candidate displayed

We found no evidence that the moral licensing effect was moderated by modern sexism (failing to support H2; see Table 1 and Supplementary Material Table S2), hostile or benevolent sexism, internal or external motivation to respond without sexism (Supplementary Material, Tables S3a–3d). Similarly, the effect of the manipulation did not depend on participants' gender, or the Democratic candidate displayed in the two conditions (Supplementary Material, Tables S4–S5b).

1.3.7. Robustness checks

The conclusions were identical when we treated the DV as a binary measure (1 = *prefer to hire a man*; 0 = *no preference for hiring a man*; Supplementary Material, Table S6). We pre-registered this analysis because the distribution of responses to similar licensing measures tends to be non-normal, with almost all participants either declining to endorse a stereotypical hiring decision or expressing a slight endorsement of this decision (see Effron et al., 2012, Study 3; Ebersole et al., 2016; online data; see overview in Supplementary Material, Table S15). As a second pre-registered robustness check, we re-ran the analyses after excluding 4% ($n = 82$) of participants who chose Trump over any of the Democrats in either the manipulation or in the end-of-study questions. The conclusions remained the same (Supplementary Material, Table S7).

2. Study 2

Study 1 suggests that endorsing a female presidential candidate did not license Democrat participants to indicate that a man was better suited than a woman for a stereotypically masculine job (i.e., no evidence for moral licensing). Study 2 aimed to address a potential methodological explanation for this effect: Perhaps Study 1's materials made participants interpret their choice of a woman as inadequate evidence of their non-sexism. We considered two versions of this explanation. First, perhaps Democrats in 2020 had such antipathy towards Donald Trump that choosing a female Democrat over Trump felt more like a repudiation of Trump than a non-sexist credential. To test this possibility, Study 2 included a condition where Democrats could endorse a female presidential candidate (e.g., Elizabeth Warren) over a male, anti-Trump Republican (e.g., Mitt Romney). Second, perhaps Democrats had such antipathy towards *Republicans* that choosing a female Democrat over Trump felt more like a repudiation of Republicans than a non-sexist credential. To test this possibility, Study 2 included a condition where Democrats could endorse a female candidate over a male, anti-Trump *independent* candidate (e.g., Angus King).

Finally, to ensure that the results could not be explained by real-

[3] Note that for one-tailed tests that go in the opposite direction than pre-registered, the one-tailed p-value is equal to 1 – (p2/2) where p2 is the two-tailed p-value.

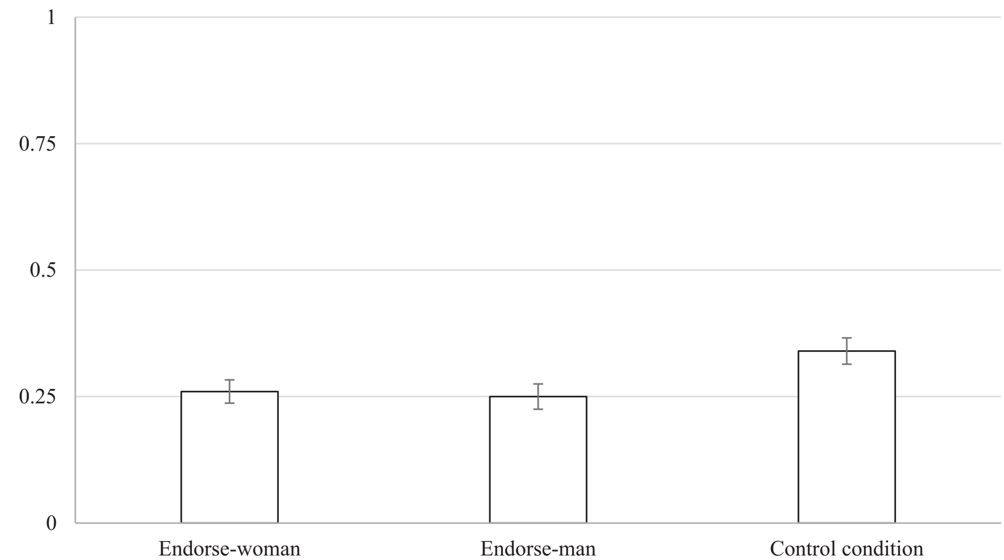

Fig. 1. Mean differences in stereotypical hiring preferences by condition (study 1).
Note. Error bars are standard errors around the mean. Stereotypical hiring preferences (the dependent measure) was assessed on a scale from −3 to +3, with positive numbers indicating a preference for hiring a man. Because −3 and + 3 were outlying scores, we replaced them with −2 and + 2 (see main text).

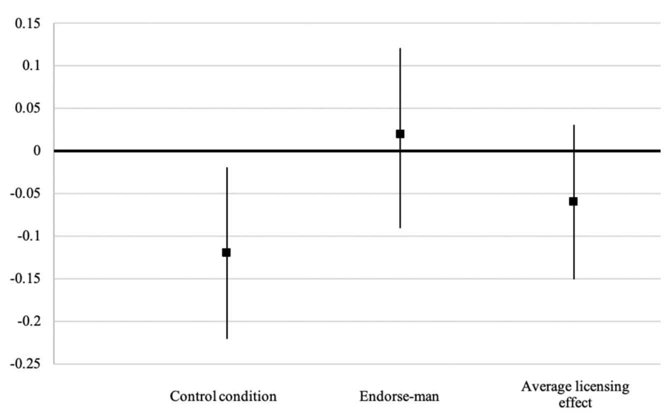

Fig. 2. Cohen's *d* by dummy variables and average licensing effect (study 1).
Note. Error bars are 95% *CIs* around Cohen's *d*. Cohen's *d* is coded so that positive numbers indicate a licensing effect. Control condition and Endorse-man were dummy coded, with Endorse-woman as the reference group. "Average licensing effect" captures the effect size of the endorse-woman condition versus the average of the control condition and the endorse-man condition.

world knowledge participants had about the specific candidates, Study 2 also included conditions where all candidates were fictional. The moral-licensing hypothesis predicts that, regardless of the specific candidates used as stimuli, endorsing a female presidential candidate would lead Democrats to express a preference for hiring men over women on a subsequent task.

2.1. Method

We pre-registered this study at aspredicted.org/blind.php?x=kv4yg4, and collected data between February 24th and March 7th 2021.

2.1.1. Participants

As in Study 1, we targeted 2400 Prolific Academic users (i.e., 800 per cell; see Study 1 for a discussion of statistical power). To recruit this sample, we used Prolific Academic's pre-screen filters to target American Democrats who voted for Joe Biden in 2020 and for Hillary Clinton in 2016, and who did not participate in Study 1. Participants were not informed of these pre-screen criteria. To improve data quality, we used filters in Qualtrics to prevent participation by people who failed a reading-comprehension question, accessed the survey on a mobile device, or resided outside the U.S.

Of the 2461 people who started the study, 2406 provided full responses, and we further excluded 178 according to our pre-registered exclusion criteria: an IP address outside the U.S., a duplicate IP

Table 1
Mean differences in stereotypical hiring preferences among conditions with modern sexism as a moderator (study 1).

Measure	Main effects model					Interactions model				
	b	SE (b)	*t*	*p*	95% *CI*	*b*	SE (b)	*t*	*p*	95% *CI*
Constant	0.26	0.03	10.34	<0.001	[0.21; 0.31]	0.25	0.03	10.19	<0.001	[0.20; 0.30]
Control condition	0.08	0.04	2.31	0.021	[0.01; 0.15]	0.10	0.04	2.71	0.007	[0.03; 0.16]
Endorse-man	−0.01	0.04	−0.17	0.868	[−0.08; 0.06]	0.00	0.04	0.02	0.984	[−0.07; 0.07]
Modern sexism						0.12	0.02	5.25	<0.001	[0.08; 0.17]
Modern sexism * Control condition						0.00	0.04	0.01	0.989	[−0.07; 0.07]
Modern sexism * Endorse-man						−0.04	0.03	−1.12	0.264	[−0.11; 0.03]

Note. $N = 2143$. The *p*-values are from two-tailed tests, but the main text reports one-tailed tests in line with our pre-registered directional predictions. Control condition and Endorse-man were dummy coded, with Endorse-woman as the reference group. Modern sexism was measured on a scale from 1 to 5. Stereotypical hiring preferences (the dependent measure) was assessed on a scale from −3 to +3, with positive numbers indicating a preference for hiring a man. Because −3 and + 3 were outlying scores, we replaced them with −2 and + 2 (see main text).

address, a duplicate Prolific ID, or saying at the end of the study that they were not a Democrat. Our final dataset contained 2228 participants (60.10% women; M_{age} = 30.88 years, SD = 10.99; see Supplementary Material, Table S8 for means, standard deviations, and correlations). At the end of the study, almost all participants confirmed their pre-screen responses that they had voted in the 2020 U.S. presidential election (95.50%) and they had or would have voted for Joe Biden (99.80%). The majority of participants further confirmed that they had voted in the 2016 U.S. presidential election (67.50%)[4] and they had or would have voted for Hillary Clinton (98.30%). A sensitivity analysis confirmed that the final sample size allowed us to detect a moral licensing effect of d = 0.14 with 85% power (α = 0.05 by a pre-registered one-tailed test).

2.1.2. Procedure

Because we ran this study after the 2020 Presidential election, we manipulated whether participants had an opportunity to endorse a woman for a 2024 presidential bid. We randomly assigned participants to one of three conditions. In the *endorse-woman-over-Republican* and *endorse-woman-over-independent* conditions (ns = 735 and 759, respectively), participants viewed a White female Democrat's name and picture next to a White male Republican's name and picture, or a White male independent's name and picture (display order randomized), and indicated whom they would vote for if those two candidates ran against each other in 2024. For each candidate, we listed information including their political party and role, and highlighted that they vocally opposed Donald Trump (e.g., "Elizabeth Warren,[5] Democrat, Senator from Massachusetts, Unsuccessfully ran for President in 2020, Vocal opponent of Donald Trump").

In the endorse-woman-over-Republican condition, we randomized the pairs to be either two real candidates (i.e., Democrat Elizabeth Warren vs. Republican Mitt Romney) or two fictitious candidates (i.e., female Democrat "Emma Wilcox" vs. male Republican "Liam Brauer"). In the endorse-woman-over-independent condition, we randomized the pairs to be either two real candidates (i.e., Elizabeth Warren vs. independent Angus King) or two fictitious candidates (i.e., "Emma Wilcox" vs. male independent "Brandon Thomas"). We ensured that the fictitious candidates' first and last names were equivalently common among White Americans.[6] In the control condition (n = 734), participants responded to the dependent measure (described below) without viewing or endorsing any candidate.

2.1.3. Measures

We administered the same measures as in Study 1 (i.e., DV: preference for hiring a man for a stereotypical masculine job in the construction industry; potential moderators: modern sexism, α = 0.83; hostile sexism, α = 0.91; benevolent sexism, α = 0.86; internal and external motivation to control sexist responding, αs = 0.78 and 0.91. respectively).

After responding to these items, participants indicated whether they had voted in the 2020 and the 2016 U.S. presidential elections. Those who said 'yes' indicated for whom they had voted, whereas those who said 'no' indicated for whom they *would* have voted. Then, participants indicated the political party they identified with or leaned towards. We used these measures to validate participants' pre-screening responses. Finally, we presented participants with the pairs of candidates they had not seen earlier in the study (in randomized order), and asked them to indicate whom they would vote for.

2.2. Results and discussion

2.2.1. Manipulation check

The majority of participants endorsed the female Democrat in the endorse-woman-over-Republican condition (96.20%) and in the endorse-woman-over-independent condition (88.90%).

2.2.2. Outliers

As pre-registered and as in Study 1, we replaced any observations that fell at least 3.29 SDs away from the grand mean (M = 0.27, SD = 0.75) with the next-smallest value. We made 76 replacements.

2.2.3. No support for the moral licensing hypothesis

As in Study 1, the results did not support the moral licensing hypothesis (see Fig. 3). Our pre-registered prediction was that giving Democrats an opportunity to endorse a female candidate over a male candidate would increase their willingness to say they would hire a man instead of a woman for the hypothetical job in the construction industry. That is, participants in the two endorse-woman conditions, compared to those in the control condition, should express a stronger preference for hiring a man.

To test this prediction, we regressed participants' hiring preference on two dummy codes for our three conditions, with the control condition as the reference group. (As pre-registered, this analysis collapsed across whether the candidates were real or fictional.) With this coding, the pre-registered moral licensing hypothesis predicts positive coefficients for each dummy code. Contrary to predictions, participants were not significantly more likely to favor men for the hypothetical job in the endorse-woman-over-Republican condition (M = 0.23, SD = 0.64) or the endorse-woman-over-independent condition (M = 0.23, SD = 0.64) than in the control condition (M = 0.31, SD = 0.67), b = −0.08, SE = 0.03, t = −2.30, p = .989, d = −0.12, 95% CI_d = [−0.22; −0.02] and b = −0.08, SE = 0.03, t = −2.38, p = .991, d = −0.12, 95% CI_d = [−0.22; 0.02], for pre-registered one-tailed tests. (Cohen's d is coded so that positive numbers indicate a licensing effect.) Note that the means are in the *opposite direction* than the moral licensing hypothesis predicts.

Comparing the control condition to the average of the two endorse-woman conditions further highlights the lack of evidence for licensing. As Fig. 4 shows, the best estimate for this comparison is an effect in the opposite direction as the licensing hypothesis, d = −0.12, and the 95% CI does not contain any values consistent with a licensing effect (i.e., positive values).

2.2.4. Exploratory analyses of condition differences

Exploratory analyses with two-tailed tests did find that the means for the three conditions differed significantly, F (2, 2225) = 3.64, p = 0.027 (see Fig. 3). Pairwise comparisons showed that participants were more likely to favor men for the job in the control condition, compared to the endorse-woman-over-Republican or the endorse-woman-over-independent condition, ts(2225) = 2.30 and 2.38, ps = 0.022 and 0.018 (two-tailed), ds both = −0.12. However, we urge caution in interpreting these effects because (a) they were in the opposite direction of our pre-registered one-tailed tests, and (b) neither effect remains significant after applying a Bonferroni correction for multiple comparisons (critical α = 0.016 after the correction).

2.2.5. Bayesian analysis

To further understand the results, we ran Bayesian regression

[4] When we did not reach our target sample size within seven days, we followed our pre-registered plan and allowed Democrats who did not vote in 2016 or who did not indicate who they voted for in 2016 to participate.

[5] We decided to focus on Elizabeth Warren instead of Vice President Kamala Harris for our Study 2 stimuli because the latter would signal not only gender and race, but also ethnicity (i.e., Harris has Indian ancestry).

[6] In a pre-test (N = 60) we asked a different sample of American Democrats who voted for Joe Biden in 2020 and for Hillary Clinton in 2016, and who did not participate in Study 1, how familiar they were with each candidate (in randomized order) on a scale from 1 (*not at all familiar*) to 7 (*very familiar*). Participants were less familiar with the fictitious candidates ($M_{\text{Emma Wilcox}}$ = 1.50, $SD_{\text{Emma Wilcox}}$ = 1.10; $M_{\text{Liam Brauer}}$ = 1.30, $SD_{\text{Liam Brauer}}$ = 0.79; $M_{\text{Brandon Thomas}}$ = 1.33, $SD_{\text{Brandon Thomas}}$ = 0.73), than with the real candidates ($M_{\text{Elizabeth Warren}}$ = 5.70, $SD_{\text{Elizabeth Warren}}$ = 1.57; $M_{\text{Mitt Romney}}$ = 5.55, $SD_{\text{Mitt Romney}}$ = 1.43; $M_{\text{Angus King}}$ = 2.07, $SD_{\text{Angus King}}$ = 1.77).

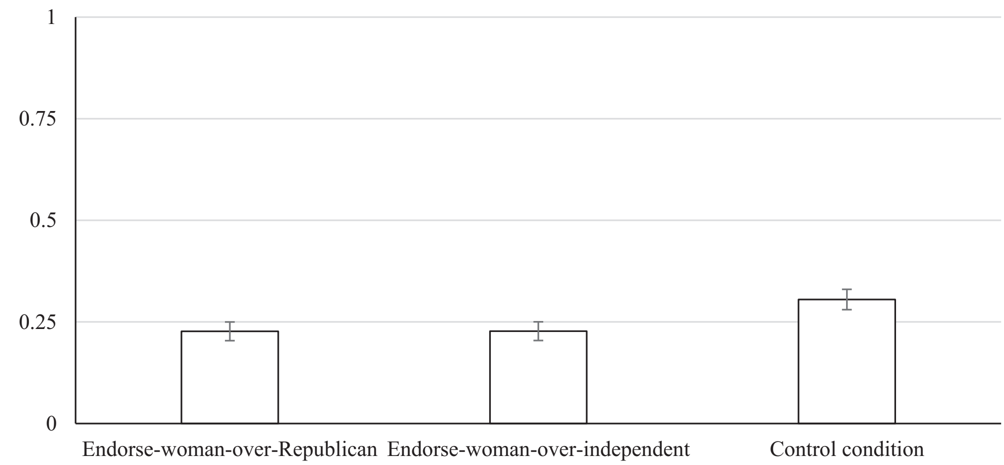

Fig. 3. Mean differences in stereotypical hiring preferences by condition (study 2).
Note. Error bars are standard errors around the mean. Stereotypical hiring preferences (the dependent measure) was assessed on a scale from −3 to +3, with positive numbers indicating a preference for hiring a man. Because −3 and + 3 were outlying scores, we replaced them with −2 and + 2 (see main text).

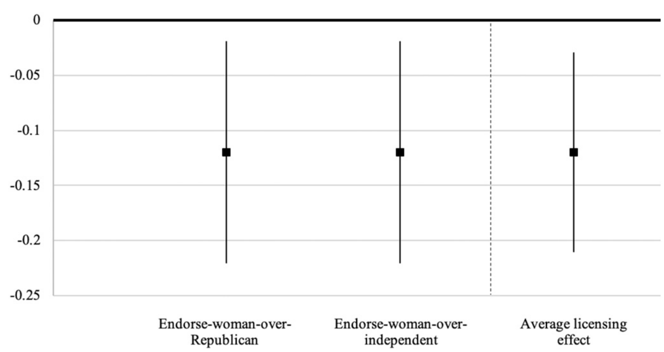

Fig. 4. Cohen's *d* by dummy variables and average licensing effect (study 2).
Note. The negative effect sizes indicate that the results were in the opposite direction than the moral licensing hypothesis. Error bars are 95% *CIs* around Cohen's *d* comparing each experimental condition to the Control condition. "Average licensing effect" captures the effect size of the Control condition versus the average of the two Endorse-woman conditions.

analysis using the procedure described in Study 1. This exploratory analysis used the same dummy codes as the pre-registered regression described above. Results suggest that the data are over four times more likely under the null model than the model with the two dummy predictors ($BF_{01} = 4.68$). Thus, the Bayesian analysis provide additional evidence that Study 2's results – like Study 1's results – are more consistent with a null effect than with a licensing effect.

2.2.6. *No moderation by gender attitudes, participant gender, or the candidate displayed*

As in Study 1, we found no evidence that the moral licensing effect was moderated by modern sexism (failing to support H2; see Table 2 and Supplementary Material Table S9), hostile or benevolent sexism, internal or external motivation to respond without sexism (Supplementary Material, Tables S10a-d). Exploratory analyses with two-tailed tests did find that the endorse-woman-over-Republican condition interacted with modern, hostile, and benevolent sexism, but not in a way that was consistent with moral licensing (see Supplementary Material for detailed results). The manipulation's effect also did not depend on participants' gender, or the Democratic candidate displayed in the two conditions (i.e., Elizabeth Warren or a fictional Democrat woman; Supplementary Material, Tables S11-S12b).

2.2.7. *Robustness checks*

The conclusions were identical when we treated the DV as a binary measure (1 = *prefer to hire a man*; 0 = *no preference for hiring a man*; Supplementary Material, Table S13). As a second pre-registered robustness check, we re-ran the analyses after excluding 11.4% (n = 255) of participants who did not endorse the female Democratic candidate in all of the pairs. The conclusions once again remained the same (Supplementary Material, Table S14).

3. General discussion

Two high-powered, pre-registered experiments found no evidence that endorsing a female presidential candidate licensed Democrats to express a more-stereotypical hiring preference. This effect emerged

Table 2
Mean differences in stereotypical hiring preferences among conditions with modern sexism as a moderator (study 2).

Measure	Main effects model					Interactions model				
	b	SE (b)	*t*	*p*	95% *CI*	*b*	SE (b)	*t*	*p*	95% *CI*
Constant	0.31	0.02	12.73	<0.001	[0.26; 0.35]	0.31	0.02	13.16	<0.001	[0.26; 0.36]
Endorse-woman-over-Republican	−0.08	0.03	−2.30	0.022	[−0.14; −0.01]	−0.08	0.03	−2.48	0.013	[−0.15; −0.02]
Endorse-woman-over-independent	−0.08	0.03	−2.38	0.018	[−0.15; −0.01]	−0.09	0.03	−2.58	0.010	[−0.15; −0.02]
Modern sexism						0.16	0.02	6.92	<0.001	[0.12; 0.21]
Modern sexism * Endorse-woman-over-Republican						−0.07	0.03	−2.00	0.046	[−0.13; −0.00]
Modern sexism * Endorse-woman-over-independent						−0.02	0.03	−0.51	0.610	[−0.08; 0.05]

Note. $N = 2228$. The *p*-values are from two-tailed tests, but the main text reports one-tailed tests in line with our pre-registered directional predictions. Endorse-woman-over-Republican and Endorse-woman-over-independent were dummy coded, with Control condition as the reference group. Modern sexism was measured on a scale from 1 to 5. Stereotypical hiring preferences (the dependent measure) was assessed on a scale from −3 to +3, with positive numbers indicating a preference for hiring a man. Because −3 and + 3 were outlying scores, we replaced them with −2 and + 2 (see main text).

regardless of participants' gender, attitudes towards women, or motivation to respond without sexism. Whereas endorsing Obama in 2008 licensed voters to favor Whites over Blacks (Effron et al., 2009), and rejecting sexist statements licensed participants to favor men over women for a hypothetical job (Monin & Miller, 2001), such moral licensing effects do not appear to generalize to Democrats' presidential politics in 2020/21.

Several considerations allow us to draw conclusions from this null result (see Maxwell, Lau, & Howard, 2015). First, with a sample size over an order of magnitude larger than almost every prior licensing study (see Blanken et al., 2015; Simbrunner & Schlegelmilch, 2017), each of our experiments provided unusually high statistical power to detect even a small licensing effect. Second, our results suggest that if a licensing effect occurred in Study 1, it would be smaller than $d = 0.03$ in Study 1 (i.e., the upper-bound of the 95% *CI* comparing the endorse-woman condition to the average of the other two conditions; see Fig. 2) – and even the smallest licensing effect would not have been consistent with the data in Study 2 (i.e., the upper-bound of the relevant 95% *CI* excluded all positive effect sizes; see Fig. 4). For comparison, the average effect size in the moral licensing literature across a wide variety of measures and manipulations is $d = 0.31$ (Blanken et al., 2015), and the effect size in a high-powered, pre-registered study using the same measure as we did was $d = 0.14$ (Ebersole et al., 2016).

Third, Bayesian analyses provided more support for a null effect than a licensing effect (see Masson, 2011). Fourth, the experiments' dependent measure should have been sensitive enough to detect a licensing effect, because the same or a similar measure has detected licensing in prior research (Bradley-Geist et al., 2010; Cascio & Plant, 2015; Ebersole et al., 2016; Effron et al., 2012; Monin & Miller, 2001), including among Democrats (Effron et al., 2009). It is possible that a more sensitive measure would have detected a gender-licensing effect, but at present the licensing literature offers no such measure. Together, these considerations minimize concerns that the present experiments failed to detect a real licensing effect (i.e., made a Type-II error). The results were more consistent with the absence rather than the presence of a licensing effect in this context.

In fact, Study 2's data were unexpectedly more aligned with a *consistency* effect than with a licensing effect. That is, a chance to endorse a woman for president subsequently made participants *less likely* to favor men over women for a stereotypically masculine job (see Fig. 4). If reliable, this finding would fit with evidence that sometimes doing good can lead people to do *more* good (e.g., Conway & Peetz, 2012; Cornelissen, Bashshur, Rode, & Le Menestrel, 2013; Mullen & Monin, 2016). However, we urge caution in interpreting this finding because it emerged from exploratory analyses and was not significant across Study 1's conditions (see Fig. 2).

Why did endorsing a woman for president not license Democrats to favor a man for a stereotypically masculine job? Study 2 allows us to rule out two salient explanations. First, perhaps participants' knowledge of the specific candidates used as stimuli in Study 1 somehow interfered with the licensing process. However, Study 2 found no evidence of licensing even when the candidates participants saw were fictional. Second, perhaps contemporary Democrats feel such antipathy towards Donald Trump in particular – or towards Republicans in general – that endorsing a female candidate in Study 1 felt more like a repudiation of Trump and his political party than like a non-sexist credential. Contrary to this possibility, however, Study 2 found no evidence that endorsing a woman over a man licensed Democrats, regardless of whether that man was an anti-Trump Republican or an anti-Trump political independent.

Having empirically addressed these two explanations, we can speculate about others. Theoretically, our paradigm would need to have met three conditions for licensing to have occurred (see Miller & Effron, 2010): Participants would need to (a) have been tempted to say on the dependent measure that the stereotypically masculine job is best suited for men, (b) have initially felt inhibited from expressing this view, and (c) have interpreted their choice of a female presidential candidate as a "non-sexist credential" that reduces this inhibition. Each criterion suggests a potential explanation for our null result.

Regarding the first criterion, it is possible that Democrats were not tempted to express a gender-stereotypical hiring preference on the dependent measure. Although research reported as recently as 2016 suggests that this measure can create such temptation (Ebersole et al., 2016), perhaps the changing conversation surrounding gender since 2017's #MeToo movement has reduced this temptation (Szekeres, Shuman, & Saguy, 2020). If this explanation were correct, however, we would still expect to see evidence of licensing among participants with higher scores on the various measures of sexism that we included in our studies. For example, participants higher in modern sexism should be more tempted to say that men are better suited than women for a stereotypically masculine job (compare to Effron et al., 2009, Study 3). However, we found no evidence of licensing even among people with higher sexism scores (see Online Supplement). Thus, it seems unlikely that insufficient temptation explains our null results.

Regarding the second criterion, it is possible that participants did not feel inhibited about expressing a preference for hiring men on the dependent measure. In other words, participants may have felt comfortable expressing this preference even without establishing non-sexist credentials, in which case our credentials manipulation would not change their behavior. The results, however, do not support this possibility either. In the control conditions – as in previous studies that successfully demonstrated licensing (see Table S15 in the Online Supplement) – only a small minority of participants said that men were better suited than women for the stereotypically masculine job (< 26% in both studies). In other words, not many people appear to have felt licensed in the control condition.

Finally, regarding the third criterion, perhaps our manipulation did not provide participants with strong enough non-sexist credentials to remove their inhibition about expressing a stereotypical hiring preference on the dependent measure. That is, participants may not have felt that endorsing a woman for president in 2020 sufficiently proved their lack of sexism. After Hillary Clinton's defeat in the 2016 presidential election, Democrats might feel that such proof would require actively promoting gender equality rather than merely stating a voting intention. Stating an intention to vote for a woman may have only signaled to Democrats a commitment to gender-equality goals; for licensing to occur, people would have to feel they had made progress towards such goals (Fishbach & Dhar, 2005).

The results of some of our exploratory analyses fit with the idea that the manipulation did not make participants feel sufficiently credentialed. Prior work suggests that when people interpret a virtuous behavior of theirs as a signal of commitment, they feel motivated to act more virtuously in the future (a consistency effect) instead of disinhibited to act less virtuously (a licensing effect; Susewind & Hoelzl, 2014; see also Mullen & Monin, 2016). Study 2's data were more aligned with a consistency effect than a licensing effect – a chance to endorse a woman for president subsequently made participants less likely to favor men over women for a stereotypically masculine job (see "average licensing effect" in Fig. 4). However, we again urge caution in interpreting this finding because it emerged from exploratory analyses and was not significant in Study 1 (see "average licensing effect" in Fig. 2).

On the one hand, we have offered some post-hoc explanations for why licensing did not occur in the present research. Of the most salient theoretical explanations, the most plausible one in our view is that Democrats in 2020/21 did not interpret their endorsement of a female presidential candidate as non-sexist credentials. On the other hand, prior research offers ample reasons to predict that licensing would occur in our experiments. A major theme of the moral licensing literature is that when people need a license, they are able to perceive seemingly trivial behaviors as adequate proof of their morality (Effron, 2016), thus making "mountains of morality from molehills of virtue" (Effron, 2014). For example, previous research found that merely stating an intention to donate blood (Cascio & Plant, 2015), agreeing to help a student in a hypothetical scenario (Khan & Dhar, 2006), or expressing disagreement with blatantly sexist statements (Monin & Miller, 2001) were all sufficient to produce a licensing effect. A priori, it is not clear that endorsing a woman for president would feel like *less* of a moral credential to participants than these other behaviors.

Thus, a key contribution of our findings is to highlight the need to sharpen theory about moral licensing to clarify which behaviors "count" as moral credentials in people's minds. Theoretically, a behavior will only "count" if people perceive it as sufficiently virtuous, interpret it as signaling progress rather than commitment to a virtuous goal, and feel that it was freely chosen (Bradley-Geist et al., 2010; Miller & Effron, 2010; Mullen & Monin, 2016). Our results illustrate, however, the challenge of predicting a priori whether a given behavior will meet these criteria. For example, do Democrats perceive endorsing a female candidate as "sufficiently virtuous," as contributing to progress towards gender equality, and as a free choice? We had assumed the answer would be yes, but our results did not support this assumption. We need more research on how people subjectively perceive their past virtuous behaviors to develop better theory about when licensing will occur (see Effron, 2014). Despite some advances on this front (e.g., Conway & Peetz, 2012; Cornelissen et al., 2013; Schwabe et al., 2018; Susewind & Hoelzl, 2014), the literature currently offers little consensus about moral licensing's boundary conditions (Mullen & Monin, 2016).

As a second contribution, our results add to a growing appreciation that moral licensing effects are not as generalizable across contexts and paradigms as initially assumed (Blanken et al., 2014; Simbrunner & Schlegelmilch, 2017; Urban et al., 2019). Robust evidence of licensing is required from high-powered experiments across multiple contexts to claim that moral licensing is a common, generalizable effect. Rigorous data do show that rejecting sexist statements can license people to subsequently favor men over women (Ebersole et al., 2016), but evidence that licensing generalizes more widely is limited by publication bias and a reliance on underpowered studies (Blanken et al., 2015). Moreover, most experimental tests of licensing use artificial laboratory manipulations (e.g., rejecting sexist statements or writing about the self using moral words; Monin & Miller, 2001; Sachdeva et al., 2009) rather than manipulating opportunities to express real-world, consequential preferences (e.g., endorsement of political candidates) – and field studies of licensing tend to be correlational (Ahmad, Klotz, & Bolino, 2020; Hofmann, Wisneski, Brandt, & Skitka, 2014; Lin, Ma, & Johnson, 2016; Schlegelmilch & Simbrunner, 2019). Thus, we need more studies like the present ones—high-powered tests of whether externally-valid manipulations can produce licensing—to understand how common and generalizable moral licensing really is.

A third contribution is that the present work helps to reduce the serious issue of publication bias that has resulted in a moral licensing literature where positive findings are over-represented (Blanken et al., 2015; Kuper & Bott, 2019; Simbrunner & Schlegelmilch, 2017). To develop a rigorous and valid moral licensing theory, the scientific record must include rigorous data about the contexts where licensing has *not* occurred—as opposed to only examples of when it *has* occurred. By reporting the null result of two high-powered studies in an impactful, real-world context, the present research begins to address this problem.

Finally, from a practical standpoint, our results address an important and timely real-world question. Prior research raised the troubling possibility that endorsing a female candidate could license people to subsequently express views that favor men over women (e.g., Effron et al., 2009). Yet, at least in our studies of American Democrats in 2020/21, this possibility did not materialize. In other words, moral licensing need not occur whenever members of an underrepresented group run for political office.

Like most other investigations of moral licensing, our conclusions are limited by the use of convenience samples and are restricted to our specific paradigm. It remains possible that a gender-licensing effect could emerge with other female political candidates in other elections, or in participant populations other than Democrats. The meaning of endorsing a political candidate from an underrepresented group will likely differ across times and contexts and future research should systematically address this possibility. Future research should also examine whether and how licensing plays out when people endorse a candidate who holds membership in more than one underrepresented group, such as Vice President Kamala Harris.

Ironically, progress towards gender equality and gender bias can go hand-in-hand. Among American Democrats in 2020/21, however, expressing support for a female candidate did not provide a license to express less gender-egalitarian views.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2021.104144.

References

Ahmad, M. G., Klotz, A. C., & Bolino, M. C. (2020). Can good followers create unethical leaders? How follower citizenship leads to leader moral licensing and unethical behavior. *The Journal of Applied Psychology*. https://doi.org/10.1037/apl0000839.

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin, 41*(4), 540–558. https://doi.org/10.1177/0146167215572134.

Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology*, 232–238. https://doi.org/10.1027/1864-9335/a000189.

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., & McKenna, C. (2010). Moral credentialing by association: The importance of choice and relationship closeness. *Personality and Social Psychology Bulletin, 36*(11), 1564–1575. https://doi.org/10.1177/0146167210385920.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Grantee Submission, 76*(1), 1–32. https://doi.org/10.1016/j.jesp.2014.09.009.

Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now? *Journal of Experimental Social Psychology, 56*, 110–116.

Chozick, A., & Parker, A. (2016). *Donald Trump's gender-based attacks on Hillary Clinton have calculated risk*. The New York Times. https://www.nytimes.com/2016/04/29/us/politics/hillary-clinton-donald-trump-women.html.

Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters, 123*(2), 191–194. https://doi.org/10.1016/j.econlet.2014.01.039.

Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin, 38*(7), 907–919. https://doi.org/10.1177/0146167212442394.

Cornelissen, G., Bashshur, M. R., Rode, J., & Le Menestrel, M. (2013). Rules or consequences? The role of ethical mind-sets in moral dynamics. *Psychological Science, 24*(4), 482–488. https://doi.org/10.1177/0956797612457376.

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin, 129*(3), 414. https://doi.org/10.1037/0033-2909.129.3.414.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., … Boucher, L. (2016). Many labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012.

Effron, D. A. (2014). Making mountains of morality from molehills of virtue: Threat causes people to overestimate their moral credentials. *Personality and Social Psychology Bulletin, 40*(8), 972–985. https://doi.org/10.1177/0146167214533131.

Effron, D. A. (2016). Beyond "being good frees us to be bad:" moral self-licensing and the fabrication of moral credentials. In *Cheating, Corruption, and Concealment: Roots of Unethical Behavior*. Cambridge, UK: Cambridge University Press. Forthcoming.

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology, 45*(3), 590–593. https://doi.org/10.1016/j.jesp.2009.02.001.

Effron, D. A., & Conway, P. (2015). When virtue leads to villainy: Advances in research on moral self-licensing. *Current Opinion in Psychology, 6*, 32–35. https://doi.org/10.1016/j.copsyc.2015.03.017.

Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103*(6), 916.

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology, 49*(3), 573–578. https://doi.org/10.1016/j.jesp.2012.08.012.

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*(2), 175–191. https://doi.org/10.3758/BF03193146.

Filipovic, J. (2017). *Our president has always degraded women—and we've always let him*. Time. https://time.com/5047771/donald-trump-comments-billy-bush/.

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research, 32*(3), 370–377. https://doi.org/10.1086/497548.

Georgeac, O., & Rattan, A. (2019). Progress in women's representation in top leadership weakens people's disturbance with gender inequality in other domains. *Journal of Experimental Psychology: General, 148*(8), 1435–1453. https://doi.org/10.1037/xge0000561.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology, 26*(2), 309–320. https://doi.org/10.1037/h0034436.

Gholamzadehmir, M., Sparks, P., & Farsides, T. (2019). Moral licensing, moral cleansing and pro-environmental behaviour: The moderating role of pro-environmental attitudes. *Journal of Environmental Psychology, 29*(1), 160–167. https://doi.org/10.1016/j.jenvp.2019.101334.

Glick, P. (2019). Gender, sexism, and the election: Did sexism help trump more than it hurt Clinton? *Politics, Groups, and Identities, 7*(3), 713–723. https://doi.org/10.1080/21565503.2019.1633931.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). Bridgesampling: An R package for estimating normalizing constants. *ArXiv Preprint ArXiv, 1710*, Article 08162.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37*(5), 701–713. https://doi.org/10.1177/0146167211400208.

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational diversity structures. *Journal of Personality and Social Psychology, 104*(3), 504–519. https://doi.org/10.1037/a0030838.

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research, 43*(2), 259–266. https://doi.org/10.1509/jmkr.43.2.259.

Klonis, S. C., Plant, E. A., & Devine, P. G. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin, 31*(9), 1237–1249. https://doi.org/10.1177/0146167205275304.

Klotz, A. C., & Bolino, M. C. (2013). Citizenship and counterproductive work behavior: A moral licensing view. *Academy of Management Review, 38*(2), 292–306. https://doi.org/10.5465/amr.2011.0109.

Kouchaki, M. (2011). Vicarious moral licensing: The influence of Others' past moral actions on moral behavior. *Journal of Personality and Social Psychology, 101*(4), 702–715. https://doi.org/10.1037/a0024552.

Kuper, N., & Bott, A. (2019). Has the evidence for moral licensing been inflated by publication bias? *Meta-Psychology, 3*. https://doi.org/10.15626/MP.2018.878.

Lin, S.-H. J., Ma, J., & Johnson, R. E. (2016). When ethical leader behavior breaks bad: How ethical leader behavior can turn abusive via Ego depletion and moral licensing. *Journal of Applied Psychology, 101*(6), 815–830. https://doi.org/10.1037/apl0000098.

List, J. A., & Momeni, F. (2020). When corporate social responsibility backfires: Evidence from a natural field experiment. *Management Science, 67*(1), 8–21. https://doi.org/10.1287/mnsc.2019.3540.

Loi, T. I., Kuhn, K. M., Sahaym, A., Butterfield, K. D., & Tripp, T. M. (2020). From helping hands to harmful acts: When and how employee volunteering promotes workplace deviance. *Journal of Applied Psychology, 105*(9), 944–958. https://doi.org/10.1037/apl0000477.

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General, 141*(1), 187–197. https://doi.org/10.1037/a0025602.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679–690. https://doi.org/10.3758/s13428-010-0049-5.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*(6), 487–498. https://doi.org/10.1037/a0039400.

Mazar, N., & Zhong, C.-B. (2010). Do green products make us better people? *Psychological Science, 21*(4), 494–498. https://doi.org/10.1177/0956797610363538.

Meijers, M. H., Verlegh, P. W., Noordewier, M. K., & Smit, E. G. (2015). The dark side of donating: How donating may license environmentally unfriendly behavior. *Social Influence, 10*(4), 250–263. https://doi.org/10.1080/15534510.2015.1092468.

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48*(3), 774–777. https://doi.org/10.1016/j.jesp.2011.12.017.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad empirical demonstrations of moral self-licensing. *Social and Personality Psychology Compass, 4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x.

Miller, D. T., & Effron, D. A. (2010). Psychological license. When it is needed and how it functions. *Advances in Experimental Social Psychology, 43*(C), 115–155. https://doi.org/10.1016/S0065-2601(10)43003-8.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*(1), 33–43. https://doi.org/10.1037/0022-3514.81.1.33.

Mullen, E., & Monin, B. (2016). Consistency vs. Licensing effects of past moral behavior. *Annual Review of Psychology, 67*, 363–385. https://doi.org/10.1146/annurev-psych-010213-115120.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006.

Ratliff, K. A., Redford, L., Conway, J., & Smith, C. T. (2019). Engendering support: Hostile sexism predicts voting for Donald Trump over Hillary Clinton in the 2016 US presidential election. *Group Processes & Intergroup Relations, 22*(4), 578–593. https://doi.org/10.1177/1368430217741203.

Rotella, A., & Barclay, P. (2020). Failure to replicate moral licensing and moral cleansing in an online experiment. *Personality and Individual Differences, 161*(109967). https://doi.org/10.1016/j.paid.2020.109967.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*(4), 523–528. https://doi.org/10.1111/j.1467-9280.2009.02326.x.

Schaffner, B. F., MacWilliams, M., & Nteta, T. (2018). Understanding white polarization in the 2016 vote for president: The sobering role of racism and sexism. *Political Science Quarterly, 133*(1), 9–34. https://doi.org/10.1002/polq.12737.

Schlegelmilch, B. B., & Simbrunner, P. (2019). Moral licensing and moral cleansing applied to company-NGO collaborations in an online context. *Journal of Business Research, 95*, 544–552. https://doi.org/10.1016/j.jbusres.2018.07.040.

Schwabe, M., Dose, D. B., & Walsh, G. (2018). Every saint has a past, and every sinner has a future: Influences of regulatory focus on Consumers' moral self-regulation. *Journal of Consumer Psychology, 28*(2), 234–252. https://doi.org/10.1002/jcpy.1025.

Simbrunner, P., & Schlegelmilch, B. B. (2017). Moral licensing: A culture-moderated meta-analysis. *Management Review Quarterly, 67*(4), 201–225. https://doi.org/10.1007/s11301-017-0128-0.

Simon, S., & O'Brien, L. T. (2015). Confronting sexism: Exploring the effect of nonsexist credentials on the costs of target confrontations. *Sex Roles, 73*(5), 245–257. https://doi.org/10.1007/s11199-015-0513-x.

Susewind, M., & Hoelzl, E. (2014). A matter of perspective: Why past moral behavior can sometimes encourage and other times discourage future moral striving. *Journal of Applied Social Psychology, 44*(3), 201–209. https://doi.org/10.1111/jasp.12214.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199.

Szekeres, H., Shuman, E., & Saguy, T. (2020). Views of sexual assault following #MeToo: The role of gender and individual differences. *Personality and Individual Differences, 166*(110203). https://doi.org/10.1016/j.paid.2020.110203.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics. 5*. Boston, MA: Pearson.

Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy, 57*, 160–171. https://doi.org/10.1016/j.enpol.2013.01.021.

Urban, J., Bahník, Š., & Kohlová, M. B. (2019). Green consumption does not make people cheat: Three attempts to replicate moral licensing effect due to pro-environmental behavior. *Journal of Environmental Psychology, 63*, 139–147. https://doi.org/10.1016/j.jenvp.2019.01.011.

Yam, K. C., Klotz, A. C., & Reynolds, S. J. (2017). From good soldiers to psychologically entitled: Examining when and why citizenship behavior leads to deviance. *Academy of Management Journal, 60*(1), 373–396. https://doi.org/10.5465/amj.2014.0234.