



UNIVERSITY OF  
PORTSMOUTH

Faculty of  
Business  
and Law

**Working Papers in Economics & Finance  
No. 2023-05**

# **Confessions of a pirate: Gender difference in survey prime to increase honest reporting**

Kate Whitman, University of Portsmouth

Zahra Murad, University of Portsmouth

Joe Cox, University of Portsmouth

Portsmouth Business School

<https://www.port.ac.uk/about-us/structure-and-governance/organisational-structure/our-academic-structure/faculty-of-business-and-law/portsmouth-business-school>

## **“Confessions of a pirate: Gender difference in survey prime to increase honest reporting”**

*Short title: Survey Primes to Increase Honesty*

### **Compliance with Ethical Standards**

- i. There are no potential conflicts of interest*
- ii. This research involves human participants. The experimental procedure and payment protocol involved transparency between researchers and participants, no deception, fair pay and respectful treatment of participants and researchers. This study was carried out in adherence to the commitments set out in the Concordat to Support Research Integrity the RCUK Policy and Guidelines on Governance of Good search Conduct.*
- iii. The participants were asked for informed consent before the start of the study. All the collected data was anonymised from the outset, hence there was no issues regarding handling of the confidentiality data.*

*This research was reviewed and approved by the Business of Law Department of the University of Portsmouth Ethics Committee in 2022*

### **Contributors:**

*Kate Whitman  
PhD Researcher  
[Kate.whitman@myport.ac.uk](mailto:Kate.whitman@myport.ac.uk)  
Tel: 07891143067*

*Dr Zahra Murad  
Senior Lecturer  
[zahra.murad@port.ac.uk](mailto:zahra.murad@port.ac.uk)  
Tel: 02392844139*

*Dr Joe Cox  
Reader in Digital Economy  
[joe.cox@port.ac.uk](mailto:joe.cox@port.ac.uk)  
Tel: 023 9284 4833*

*University of Portsmouth  
School of Accounting, Economics and Finance  
University of Portsmouth  
Portsmouth PO1 3DE  
United Kingdom*

# Confessions of a pirate: Gender difference in survey prime to increase honest reporting

## Abstract

Survey data is essential for marketing and scientific research. However, recent evidence suggests that men and women may underreport undesirable behavior to different degrees and for different motivations, making it difficult for marketers to trust consumer data. Two survey experiments were conducted to test priming effects aimed at minimizing social desirability bias, hypothesizing a gender difference in efficacy. Using digital piracy as an example of an underreported behavior, Study 1 shows that a positive cues condition, which is designed to provide respondents with convenient rationalizations, increases undesirable behavior reporting. Negative primes have a greater inhibitory effect on men's reporting of undesirable behavior compared to women's, thus reversing the gender reporting gap. Study 2 explores the relationship between measured social desirability bias, positive cues, and gender. We find that the treatment has the strongest effect on men and only significantly affects participants with high social desirability bias. When considering both studies (N = 1,734) we estimate that the positive cues treatment increases the amount of piracy participants are willing to report by 42%. Market researchers are recommended to add positive cues before questions about undesirable behavior, especially in the case of men. Furthermore, sequential undesirable behavior questions are likely to increasingly inhibit men's reporting, suggesting that market researchers should randomize these sensitive questions.

## Introduction

Evidence suggests that social desirability bias, the tendency for survey respondents to downplay undesirable behaviors or exaggerate desirable behaviors to appear more socially acceptable (Marlowe & Crowne, 1960), poses a threat to the validity of market research. A broad range of research topics perceived as socially sensitive are reported to be at risk from social desirability bias including green behaviors, unhealthy habits, and political persuasion (Sudman & Bradburn; 1978, Kim & Kim, 2016; Holbrook & Krosnick, 2010; Davis, Thake & Vilhena, 2010; Larson & Kinsey, 2019; Klar, Weber & Krupnikov, 2016). A growing body of literature indicates that social desirability bias is gendered, with men and women underreporting undesirable behaviors to different extents and for different motivations (Bernardi, 2006; Rosenman, Tennekoon & Hill, 2011; Chung & Monroe, 2003; Dalton & Ortegren, 2011; Lalwani, Lee, Shrum, & Viswanathan, 2023). This indicates further problems for marketers relying on survey data, as one gender's attitudes or behavior may be misrepresented relative to the other.

Over two studies this paper investigates the influence of survey primes on social desirability bias, measured by men's and women's willingness to report undesirable behavior. Survey priming is a survey or question design or wording manipulation used by marketers and psychologists to influence respondents' answers (Sintov & Prescott, 2011). The literature on using priming effects to enhance honest reporting of sensitive topics presents mixed findings. The broader survey design literature assumes that diffusing the emphasis on questions about undesirable behavior by interspersing them with neutral or positive behavior questions reduces participants' perceived threat (Krumpal, 2013) and "*diminishes the focus on a specific behavior question*" (Sudman and Bradburn, 1982, p. 61). Survey design guides, like Vinten (1995), also advocate for this approach. However, as highlighted by Tourangeau and Yan

(2007), these recommendations lack empirical evidence. Forgiving and permissive language primes have produced contradictory results, with some studies suggesting that these primes increase honest reporting (Charles & Dattalo, 2018), and others finding no effect (Abelson, Loftus & Greenwald, 1992).

A recent paper by Lalwani, Lee, Shrum, and Viswanathan (2023) indicates that men and women tend to bias their answers for different motivations: men to promote and women to protect their image. If men and women bias their responses to sensitive topics to different degrees and for different motivations, it is plausible that survey primes may influence them differently. However, little research has been conducted to test this. If a gender difference exists, it may explain why the literature on the efficacy of survey primes designed to enhance honest reporting is so contradictory. Therefore, over two survey experiments on a random sample, four survey primes are tested for their effect on undesirable behavior reporting. Positive and negative attitude primes (permissive and restrictive) are compared to positive and negative behavior primes (positive cues and contrast). The results suggest that behavior primes have a greater impact on respondents' willingness to report undesirable behavior than attitude primes. However, both negative attitude and negative behavior primes inhibit men's honest responses.

The positive attitude prime (permissive) has no impact on men's reporting and serves to inhibit women's reporting. However, the positive behavior prime, *positive cues*, prompts participants to report positively framed common behaviors before answering the target question. This prime successfully increases the reported incidence of undesirable behavior compared to a control group with no prime. Study 1 indicates that the effect is greater for men. Study 2 examines the influence of the positive cues prime in more depth, finding that it is most effective for men and both genders with a high social desirability bias score. When considering both genders together, the use of positive cues increased reporting by 42% across the two studies in this paper (N = 1,734). The positive cues method can be easily applied by marketers to improve market research validity, requiring the addition of a few simple preceding questions to a sensitive topic. The positive cues method has been successfully used to increase green behaviors and financial planning (Cornelissen et al., 2008; Hoffmann & Plotkina, 2021), and our study contributes to this literature by presenting it as a novel method of reducing social desirability bias.

## Theoretical Framework

Näher and Krumpal (2012) proposed that social desirability bias, driven by cognitive dissonance resulting from conflicting attitudes and behavior, can be reduced through careful survey design. Research indicates that people prefer to avoid cognitive dissonance and strive for consistency with their past behavior (Festinger, 1957) and self-perception (Bem, 1972). However, acting morally may not always align with an individual's self-interest (e.g., avoiding paying tax or buying green but more expensive products). People may employ strategies to minimize the negative impact of acting against their values, such as changing cognitions, adding rationalizations, or reducing conflicting cognitions (Festinger, 1957; Fukukawa, Zaharie, & Romoñi-Maniu, 2019). This paper theorizes that priming cognitive dissonance reduction strategies will enhance honest self-reporting by decreasing the social desirability of undesirable behaviors

Studies on ethical behavior differences between genders often find that women exhibit higher ethical behavior and consumption (Glover, Bumpus, Sharp & Munchus, 2002; Luchs & Mooradian, 2012; Higgins, 2006; Morris, Johnson & Higgins, 2009). Gender differences in ethical behavior have been attributed to variations in sensitivity to social norms, social groups, experiences of the injunctive norm, and social identity (Chung & Monroe, 2003; Akers, 2011; Tajfel & Turner, 1985). These differences in social influence suggest a potential confounding

variable in social desirability bias. If women face more social pressure to behave ethically, they may also feel more pressure to report ethically, which can deflate the frequency or intensity of their self-reported unethical behavior. Supporting this, contemporary literature suggests that women are more likely than men to bias their responses regarding undesirable behaviors (Rosenman, Tennekoon & Hill, 2011; Bernardi, 2006). Additionally, Dalton and Ortegren (2011) found that controlling for social desirability bias reduces the gender difference in reporting various unethical behaviors.

Evidence suggests that men and women employ different mechanisms to overcome cognitive dissonance and social desirability bias. In a study on meat consumption, Piazza et al. (2015) found that men are more inclined to use justifications to neutralize their unethical behavior. Similarly, Riekkinen (2016) discovered that men are more likely to agree with rationalizations for undesirable behavior. Rothgerber (2013) observed that men tend to use direct, unapologetic strategies such as defensive justifications to cope with cognitive dissonance. On the other hand, women employ indirect, apologetic strategies such as avoidance and underreporting. Lalwani, Lee, Shrum, and Viswanathan (2023) found that men engage in self-deceptive, promoting strategies in survey reporting, while women employ impression management and protection strategies. This suggests that women may be more sensitive to primes that threaten their self-image, which could reduce their willingness to report undesirable behaviors. With these findings in mind, Study 1 examines gender differences in survey contexts designed to increase or reduce cognitive dissonance and social desirability bias, measured by participants' willingness to report undesirable behavior.

#### Priming Positive vs Negative Attitudes: “Permissive” and “Restrictive” Treatments

Näher and Krumpal (2012) propose that survey primes framing undesirable or unethical behavior in a forgiving or permissive manner reduce cognitive dissonance by minimizing the conflicting cognition's intensity during reporting. Similarly, shaming or restrictive contexts should amplify cognitive dissonance and reporting bias. Catania et al. (1996) found that supportively worded questions increased response rates and suggested the presence of a "judgmental factor" contributing to social desirability bias. Creating a positive context for socially undesirable subjects is often recommended in survey design literature to elicit more honest responses (Sudman & Bradburn, 1982). However, the literature on using forgiving language to reduce social desirability bias presents inconsistent findings. Charles and Dattalo (2018) found that forgiving language decreased social desirability bias, while Abelson, Loftus, and Greenwald (1992) found no significant effect. Tourangeau and Smith (1996) conducted an experiment on the number of sexual partners individuals were willing to report, phrasing preceding questions restrictively or permissively for both men and women. Contrary to expectations, the restrictive questions yielded more honest responses than the permissive questions. Our study contributes to the literature by conducting a more controlled examination of these treatments, using the same wording for men and women (assuming both are more likely to underreport than overreport) and presenting the same examples for permissive and restrictive questions, only changing the positive or negative framing. Given that women are suggested to be more motivated by social desirability (impression management) than men (Lalwani et al., 2023), we hypothesize that the judgmental factor will have a greater impact on women's reporting bias.

*Hypothesis 1a:* Participants in the permissive group will report more undesirable behavior compared to the control group with no context.

*Hypothesis 1b:* The effect will be more pronounced among women than men.

*Hypothesis 2a:* Participants in the restrictive group will report less undesirable behavior compared to the control group with no context.

*Hypothesis 2a:* The effect will be more pronounced among women than men.

### Priming Positive Behaviors: “Positive Cues” Treatment

Steele and Liu (1983) demonstrated that cognitive dissonance can be alleviated by providing alternative means to enhance one's self-image. We propose that allowing participants to report offsetting good behaviors through positive cues will increase their likelihood of reporting undesirable behavior. Cornelissen et al. (2008) employed positive cueing to enhance moral self-perception, good intentions, and good behavior. This method involves framing common behaviors positively and allowing participants to report them. Positive cueing influences a preference for moral consistency, as participants' past behaviors are framed as positive, leading them to be more inclined to exhibit positive behavior in the future. Hoffmann and Plotkina (2021) used positive framing of financial circumstances to encourage better retirement planning, suggesting that improved self-perception enhances self-efficacy. However, the effectiveness of positive cues in reducing social desirability bias has not been tested. Elevating moral identity by positively framing past behaviors may lead participants to hide their unethical behavior to maintain their elevated identity. On the other hand, highlighting related positive behaviors might provide the necessary compensatory cognitions to rationalize their behavior. Although not specifically testing positive cueing, Duff et al. (2007) found that providing participants with excuses not to vote improved honest reporting. Thus, we compare two opposing hypotheses to uncover the mechanisms behind the impact of positive cues on honest reporting. Studies like Piazza et al. (2015) suggest that men are more prone to agreeing with rationalizations, and Lalwani et al. (2023) propose that men are more motivated by self-promotion. Therefore, we hypothesize that a prime that allows participants to boast about their past good behavior and emphasizes rationales will have a stronger effect on men than women.

*Hypothesis 3a:* Participants in the positive cues group will report less undesirable behavior compared to the control group with no context.

*Hypothesis 3b:* Participants in the positive cues group will report more undesirable behavior compared to the control group with no context.

*Hypothesis 3c:* Men will be more influenced by the positive cues treatment than women.

### Priming Negative Behaviors: “Contrast Effect” Treatment

Another psychological mechanism suspected to influence reporting is the contrast effect. Kalton, Collins, and Brook (1978) suggest that desensitization to behaviors can occur when they are presented as more acceptable. One way to make a behavior appear more acceptable is to contrast it with worse behavior. Schwarz and Bless (1992) conducted a study where participants rated the trustworthiness of a politician. In one group, participants first rated a politician involved in a scandal, and then they rated a non-scandalized politician's trustworthiness. The participants rated the non-scandalized politician as more trustworthy compared to those who were not asked about the scandal. The researchers suggest that in contrast to the scandalized politician, the non-scandalized one seemed relatively honest. However, there is limited evidence of contrast effects in reducing social desirability bias. Gendall, Hoek, and Blakeley (1992) experimented to investigate how survey design influences the admission of drunk driving offenses. They found that adding more extreme and socially undesirable driving offenses to the response options made people more likely to admit to drunk driving. This suggests that the perceived risk of reporting an undesirable activity may decrease when the activity is compared to worse behaviors. We theorize that this occurs because cognitive dissonance can be reduced by minimizing the conflicting cognition related to the target behavior. Research suggests self-standards are not static and fluctuate depending on the social context (Swann & Schroeder, 1995; Stone & Cooper, 2001). We suggest that asking respondents about more negative behaviors before asking them about the target behavior will make the target behavior seem relatively less negative, thereby reducing social desirability bias. As Dalton and Ortegren (2011) suggest that more women than men underreport due to

impression management, we hypothesize that decreasing the relative impact of undesirable behavior by comparing it to more extreme behaviors will have a greater effect on women.

*Hypothesis 4a:* Participants in the contrast group will report more undesirable behavior compared to the control group with no context.

*Hypothesis 4b:* The effect will be more pronounced among women than among men.

#### Neutral Context

Kingston and Dorans (1984) found that the location of survey items influences survey reporting, and order effects are well-researched phenomena that impact self-reporting (Krosnick & Alwin, 1987). To account for the potential effects of placing a target question after any context, a neutral context was included in the experiment. Participants were asked to answer four morally neutral questions as a placebo condition before the target question. Since the neutral context should not alter the moral context, we do not expect it to differ significantly from the control group. Therefore, the preceding treatments will be compared to both the control group and the neutral group as a robustness measure.

*Hypothesis 5:* Participants in the neutral group will report similar levels of undesirable behaviour to the control group with no context.

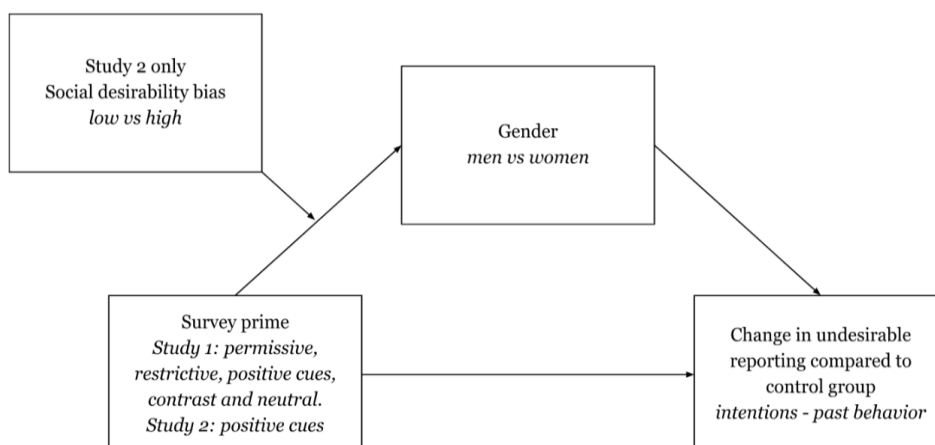
King and Bruner (2000) suggest that social desirability bias is the most common type of response bias and recommend testing for social desirability bias to reduce data contamination. Chung and Monroe (2003) propose that the more unethical the behavior, the more likely people are to bias their responses. Näher and Krumpal (2012) posit that this relationship occurs because reporting unethical behaviors that contradict personal or social values triggers cognitive dissonance, leading to biased responses. Therefore, Study 2 explicitly measures respondents' susceptibility to social desirability bias to investigate whether it explains the effect of the survey prime.

Across various countries, Bernardi (2006) found that women scored significantly higher on social desirability scales than men, calling into question whether women behave more ethically than men or simply answer surveys more desirably. Dalton and Ortegren (2011) extended this line of research finding that over 30 different ethical questions, gender-different responses are greatly reduced when social desirability bias is controlled for. Therefore, by using a similar measure, the survey prime treatment can be specifically targeted at respondents with high social desirability bias. By assuming that respondents with a low social desirability bias score are answering fairly honestly across genders, we would expect to see the truest gender response on respondents with a high social desirability bias. This leads us to test the following hypothesis:

*Hypothesis 7:* Social desirability bias will moderate the gender effect of the treatment.

The conceptual model and research hypotheses for both Study 1 and 2 can be seen in *Figure 1*.

Figure 1: Conceptual model and research hypotheses



## Overview of Studies

In this paper, we focus on digital piracy as an example of undesirable behavior. Digital piracy refers to the illegal copying or downloading of copyrighted software and media files (Al-Rafee & Cronan, 2006, p. 237). Given the private nature of digital piracy, researchers, marketers, and industry experts heavily rely on survey data. However, social desirability bias has been identified as a threat to the validity of piracy data (Phau, Teah & Liang, 2016; Cheng, Sims & Teegen, 1997; Gergely & Rao, 2014; Christensen & Eining, 1991). Therefore, it is crucial to consider and mitigate social desirability bias when studying self-reported digital piracy data.

Two studies examine the interplay between survey priming effects and gender, with Study 2 including a social desirability bias explanatory measure. An online sample of participants aged 18 - 80 was recruited through Prolific (prolific.co) and paid the equivalent of £8 per hour. After reading a short background section, participants were randomly assigned into treatment groups with an equal gender split. Each study featured a control group with no preceding priming section before the target question (the dependent variable). This was compared to treatment groups which had priming questions before the target question. The target question is designed to measure how much digital piracy respondents are willing to report, assuming that participants are generally unlikely to overreport (as piracy is illegal). The statement was, "I stream or download infringing content (TV, music, etc.) from a digital piracy site (torrent, cyberlocker, stream-rippers, etc.) \_\_\_\_\_ times a week. *For the purposes of this question, 1 = 1 download (regardless of file size), and 1 = session of streamed content (regardless of length of stream).*" Participants who failed the attention check were removed from the analysis.

## Study 1

### Method

A total of 1200 participants were recruited for the study, with 200 participants randomly assigned to each treatment group. The respondents were recruited through Prolific (prolific.co). The sample size was determined to detect a small effect size of  $f^2 = 0.05$  in a multiple linear regression model, with an error probability of 0.05 and a power of 80%. As gender is another variable for analysis, the final sample size was 98 participants per group. Research suggests that gender differences in memory (Baer, Trumpeter, & Weathington, 2006) and recall bias (Beyer, 1998) may exist. Therefore, half of the sample (50%) were asked



to report the amount of piracy (number of illegal downloads or streams) accessed in the last week, while the other half were asked to report the amount of piracy accessed in the last three months, with the answer divided by 13 to make it equivalent to one week. The one-week data aimed to minimize memory issues, while the three-month data aimed to reduce the number of zeros in the data. In our analysis, we controlled for the two different measurements by including the DV week/three-month dummy in linear regression (see Appendix B). After removing participants who failed the attention check, a final sample of  $N=1161$  remained. All respondents provided basic demographic information and were either given four priming questions followed by the target question or, in the case of the control group, asked only the target question without any priming questions. Specific question details for each treatment can be found in *Table 1* and are described below.

*Table 1: Details of treatment groups used in Study 1*

<b>Condition</b>	<b>Questions (answered by a 5-point Likert)</b>
Positive Attitudes “Permissive”	Digital piracy is sometimes acceptable Digital piracy does no real harm to the careers of individual creatives If digital content were more affordable, digital piracy wouldn’t be necessary There should be no punishment for digital piracy
Negative Attitudes “Restrictive”	Digital piracy is never acceptable Digital piracy ruins the careers of individual creatives Even if legitimate content could be cheaper, digital piracy is theft There should be punishments for digital piracy
Positive Behaviors “Positive Cue”	I support the arts by purchasing tickets for the cinema or live events I support the arts by purchasing legal TV/music subscriptions (such as Netflix or Spotify) I support creative talent by sharing an artist’s film/music release on my social media page I help promote creative content by recommending a film/music release to people I know
Negative Behaviors “Contrast Effect”	I regularly take cash-in-hand jobs to avoid paying taxes I sometimes produce counterfeit goods to sell online I often call in sick from work despite being healthy and fit for work I regularly, knowingly buy counterfeit or stolen items, either online or offline
Neutral	I prefer rock music to pop music My favorite film genre is romantic comedy I prefer to watch films at home than at the cinema My favorite band is The Beatles

## Results

*Table 2* indicates five models of multiple linear regressions examining the effect of the treatments compared to the control group on the amount of piracy reported. A positive coefficient indicates that more piracy is reported in the treatment than in the control group, suggesting that the treatment increases honest reporting, as respondents have no incentive to inflate their unethical behavior. Model 1 indicates that men report significantly more piracy than women ( $p\text{-value} < 0.01$ ). Model 2 which includes gender as a control variable suggests that positive cues significantly increase honest reporting ( $p\text{-value} = 0.04$ ) and the contrast treatment significantly reduces honest reporting ( $p\text{-value} = 0.05$ ). Model 3 indicates an interaction effect between gender and the restrictive ( $p\text{-value} = 0.02$ ) and contrast ( $p\text{-value} = 0.09$ ) groups. Models 4 and 5 indicate that men report less piracy in the restrictive ( $p\text{-value} = 0.02$ ) and contrast treatments ( $p\text{-value} = 0.04$ ) and women report less piracy in the permissive treatment ( $p\text{-value} = 0.05$ ), compared to the control group. The interaction effect between positive cues and gender is non-significant ( $p\text{-value} = 0.48$ ). However, when looking at the

genders separately in models 4 and 5, the positive cues treatment has a weakly significant positive coefficient for men (p-value = 0.07), and a non-significant positive coefficient for women (p-value = 0.29).

Table 2: Linear regressions of treatment group vs control group including gender effects

Model	1	2	3	4	5	6
				Male	Female	All
Male	0.564*** (0.164)	0.558*** (0.164)	0.663** (0.286)			
Permissive		-0.267 (0.190)	-0.243** (0.122)	-0.291 (0.349)	-0.243** (0.122)	-0.257 (0.192)
Restrictive		-0.234 (0.202)	0.234 (0.276)	-0.695** (0.288)	0.234 (0.276)	-0.237 (0.202)
Positive Cues		0.817** (0.393)	0.532 (0.499)	1.086* (0.603)	0.532 (0.499)	0.804** (0.392)
Contrast		-0.314* (0.165)	-0.046 (0.156)	-0.604** (0.288)	-0.046 (0.156)	-0.329** (0.166)
Neutral		0.087 (0.263)	-0.080 (0.160)	0.289 (0.518)	-0.080 (0.160)	0.060 (0.262)
Permissive x Male			-0.048 (0.369)			
Restrictive x Male			-0.929** (0.399)			
Positive Cues x Male			0.554 (0.783)			
Contrast x Male			-0.558* (0.328)			
Neutral x Male			0.369 (0.542)			
Constant	0.417*** (0.095)	0.408*** (0.152)	0.355*** (0.114)	1.018*** (0.262)	0.355*** (0.114)	0.692*** (0.146)
Observations	1,156	1,156	1,156	577	579	1,161
Adjusted R-squared	0.009	0.024	0.028	0.027	0.003	0.014

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

### Robustness Analysis

As a robustness analysis, the treatment groups were compared to the neutral condition in linear regression (see Appendix B). Though with weaker statistical significance than when compared to the control group, the results indicate that the positive cues condition influenced the participants to report more piracy (p-value = 0.08) and the contrast treatment influenced them to report less piracy (p-value = 0.09), compared to the neutral condition. There was no difference between reported piracy between the control and the neutral conditions (p-value = 0.82). Furthermore, interaction effects between gender and the contrast (p-value = 0.06) and restrictive (p-value = 0.02) treatments can still be seen – with these treatments inhibiting

men's reporting. These results indicate that the treatments had an influence because of the manipulation, rather than a placebo effect of *any* four preceding questions.

Secondly, a linear regression compared the groups by the timeframe of the dependent variable (one week vs three months). This analysis halved the sample size so the results should be interpreted cautiously. The positive cues treatment influenced more piracy reporting than the control group in both timeframes, though was non-significant in the one-week time frame (p-value = 0.20) and was significant in the three-month time frame (p-value = 0.05). As the sample size further decreased when we analyze gender interaction effects, no statistically significant effects were detected, however, the coefficients had qualitatively the same signs as in the main regression analysis.

### Study 1 Discussion

Gender differences in survey priming effects were compared to a control group where a target question was asked directly. The control group results showed that men reported significantly higher levels of piracy compared to women. The "positive cues" treatment, which involved asking respondents to report common behaviors framed in a morally-positive manner before the piracy question, led to increased honest reporting, providing support for hypothesis 3b. This treatment aimed to rationalize piracy by allowing respondents to first report behaviors that support the arts. Based on Riekkinen's (2016) suggestion that men are more likely than women to use rationalizations to overcome cognitive dissonance, we hypothesized (3c) that the treatment would be more effective for men. Our findings revealed that the treatment increased piracy reporting for men, although not to a statistically significant extent compared to women's reporting.

The other treatments showed either no significant difference in the reported amount of piracy or resulted in a reduction, indicating the importance for marketers and academics to carefully construct surveys. The contrast effect treatment, which primed respondents with more extremely negative behaviors, was hypothesized to make piracy appear relatively less threatening, especially for women. However, contrary to hypothesis 4a and the findings of Gendall, Hoek, and Blakeley (1992), this treatment led to a reduction in piracy reporting, particularly among men, closing the gender reporting gap. It's possible that the behaviors used in the experiment were not extreme enough to make piracy seem innocuous in comparison. Alternatively, mentioning illegal activities may have made respondents feel threatened, causing them to be more guarded and underreport their piracy. This finding may be of interest to researchers examining multiple undesirable behaviors in a single survey. Similarly, the restrictive condition, with negatively framed attitude questions, had a stronger negative impact on men's piracy reporting compared to women's, reversing the gender reporting gap. Contrary to hypothesis 4b, the results suggest that men are more sensitive to the judgmental factor than women. Furthermore, the permissive condition had minimal influence and even a slight negative effect on women's reporting, contradicting hypotheses 1a and 1b. Therefore, although negative attitude primes can inhibit honest reporting, their positive alternatives do not seem to have the opposite effect. This supports the findings of Tourangeau and Smith (1996), who found that permissive and restrictive contexts were ineffective in influencing social desirability bias in the desired direction.

In general, the treatments had a stronger influence on men's reporting compared to women's, whether positively or negatively. The reporting of behaviors (positive cues and contrast treatments) had a greater impact than the reporting of attitudes (permissive and restrictive treatments) in altering the willingness to report unethical behavior. Moreover, the results suggest that answering positive behavior questions before a negative behavior increases reporting while answering negative behavior questions before negative behavior decreases reporting, although a larger sample size is needed to make definitive claims. This implies that the perceived threat of sensitive questions can be reduced by emphasizing good behavior or

increased by highlighting bad behavior. Further research is necessary to thoroughly examine these experimental treatments, including the identification of moderating factors.

The main contribution of Study 1 to the literature is providing evidence to support the idea that positive cues increase piracy reporting, confirming hypothesis 3a. This finding has implications for interpreting previous survey-based piracy data and, more broadly, for designing surveys on undesirable behavior. However, the study did not directly include explanatory variables related to social desirability. As the focus of this paper is to identify primes that enhance honest reporting, Study 2 was designed to test whether the effect of positive cues on a larger sample size would still increase honest reporting, exhibit a more pronounced gender effect, and explore the role of social desirability bias as an explanatory factor.

## Study 2

Study 1 demonstrated that the positive cues treatment, which involved participants answering four commonly reported and positively framed questions before the target question, resulted in increased piracy reporting. We hypothesized that this would occur because the positive behaviors would provide a convenient rationalization for the undesirable behavior in the target question. For example, by asking participants whether they have a legal subscription to an entertainment service or visit the cinema, behaviors framed to support the arts, the participant would feel more entitled to engage in a behavior that is harmful to the arts, i.e. piracy, decreasing their social desirability bias. We found that the effect was stronger for men, though the interaction effect between gender and positive cues was not significant. Study 2 aims to investigate whether controlling for social desirability bias, which is indicated to be gendered, would moderate the gender effect seen in Study 1.

### Method

In Study 2, the method used in Study 1 was replicated, but with two treatments: the control group and the positive cues treatment. This allowed for a larger sample size in the treatment groups, resulting in a total of 578 participants after excluding those who failed the attention check.

Before conducting Study 2, a pilot study (N = 100) was conducted to test a larger set of common, positive behaviors, which can be found in Appendix A. Nine behaviors were selected from art websites (lightspacetime.art, theartling.com). The pilot study assessed participants' ability to readily report the suggested behaviors (as they were common), agreement on the behaviors supporting the arts as suggested in the question wording, and whether the statements made them feel ethically positive. Five of the behaviors were commonly reported, and participants agreed that they supported the arts and evoked ethical positivity. These five behaviors were similar in nature to the four statements used in Study 1, further validating the treatment manipulation in that study.

The relationship between the treatment group and the dependent variable (reported amount of piracy) was analyzed using regression, considering the factors of gender and the participants' proneness to social desirability bias. Social desirability bias was measured using the 20-item Impression Management Scale developed by Paulhus (1991). This scale is considered a reliable measure and has been recommended by Dalton and Ortegren (2011) for comparing gender differences in proneness to social desirability bias. To create individual-level participant variables for high and low social desirability bias groups, the Impression Management Scale scores were binarized by using the median as the cutoff point.

## Results

Non-parametric Wilcoxon rank-sum tests were conducted to compare the high vs low social desirability bias piracy reported in the control group: this was statistically significant ( $p$ -value = 0.01), with the high social desirability bias group reporting less piracy ( $M = 2.22$ ,  $SD = 6.01$ ) than the low social desirability bias group ( $M = 3.95$ ,  $SD = 14.85$ ). When comparing social desirability bias groups and treatment groups, we found that there was no statistical difference in piracy reporting between the control group and the positive cues group for the low social desirability bias group ( $p$ -value = 0.35). However, there was a statistical difference between the control group and the positive cues group for the high social desirability bias group participants ( $p$ -value = 0.01).

A linear regression analysis was conducted with interaction effects between gender, the positive cues treatment, and social desirability bias (see *Table 3*). Models 1-3 test for hypothesis 6, while models 4-6 test for hypothesis 7. Models 1 and 2 demonstrate that the positive cues treatment did not yield a significant increase in the reported amount of piracy among the low social desirability bias group ( $p$ -value = 0.94). However, it was effective in increasing reporting among the high social desirability bias group ( $p$ -value = 0.02). The interaction between social desirability bias and positive cues (model 3) had a positive coefficient, indicating that people in the high social desirability bias group were more likely to increase their reported piracy as a result of the positive cues treatment, while this was not the case in the low social desirability bias group. Figure C2 in Appendix C provides a visual representation of this finding. However, the interaction effect is not statistically significant ( $p$ -value = 0.38). This non-significant  $p$ -value is a result of the large confidence intervals in the low social desirability bias group due to the nature of the data, however, the effect is in the expected direction.

Models 4 and 5 examined the interaction between social desirability bias and positive cues separately for men and women. The results indicate that the interaction effect was not significant for men ( $p$ -value = 0.94), suggesting that social desirability bias did not moderate the effect of positive cues on piracy reporting among men. However, for women, the interaction effect was significant ( $p$ -value = 0.07), indicating that social desirability bias did moderate the effect of positive cues on piracy reporting for women. Model 6 explored the three-way interaction between gender, social desirability bias, and positive cues. The results revealed that this three-way interaction was not statistically significant ( $p$ -value = 0.26). Specifically, for men, positive cues improved piracy reporting regardless of their social desirability bias group. However, for women, positive cues only had a significant effect on improving honest reporting among those in the high social desirability bias group, while the reverse was the case in the low social desirability bias group. Figure C3 in Appendix C visually illustrates these findings. We can observe large confidence intervals in the low social desirability bias group, which contributed to the non-statistical significance of the triple interaction term, however, the effect is in the expected direction.

Table 3: Linear regressions looking at the effect of gender, social desirability bias (SDB), and treatment group

Model	1 Low SDB	2 High SDB	3 All SDB	4 Male	5 Female	6 All Genders
Positive Cues	-0.123 (1.585)	1.344** (0.566)	-0.123 (1.585)	2.055 (2.458)	-2.570 (1.894)	-2.570 (1.894)
SDB			-3.200*** (1.060)	-2.666** (1.301)	-3.582** (1.755)	-3.582** (1.755)
Positive Cues x SDB			1.467 (1.683)	-0.221 (2.684)	3.508* (1.959)	3.508* (1.959)
Male						0.510 (2.120)
Positive Cues x Male						4.624 (3.103)
SDB x Male						0.917 (2.184)
SDB x Male x Positive Cues						-3.729 (3.323)
Constant	4.077*** (1.034)	0.877*** (0.234)	4.077*** (1.034)	4.316*** (1.193)	3.806** (1.753)	3.806** (1.754)
Observations	296	282	578	279	298	577
Adjusted R-squared	0.003	0.017	0.011	0.008	0.018	0.025

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

### Study 2 Discussion

In summary, Study 2 established that the positive cues treatment effectively increased piracy reporting among participants with high proneness to social desirability bias, supporting hypothesis 6 (survey primes will increase honest reporting for respondents in the high social desirability bias group). However, the treatment did not have a significant effect on participants with low proneness to social desirability bias. This was partly due to low social desirability bias participants providing extremely noisy reports, which could have been tested with a much larger sample size. However, the power calculation with a 5% error probability, 80% power, and small effect sizes would require around 800 observations per treatment group to achieve statistical significance, which in this study was not accounted for and would only detect very small effect sizes. We thus relied on the qualitative and directional interpretation of the coefficients when the standard deviations were too high.

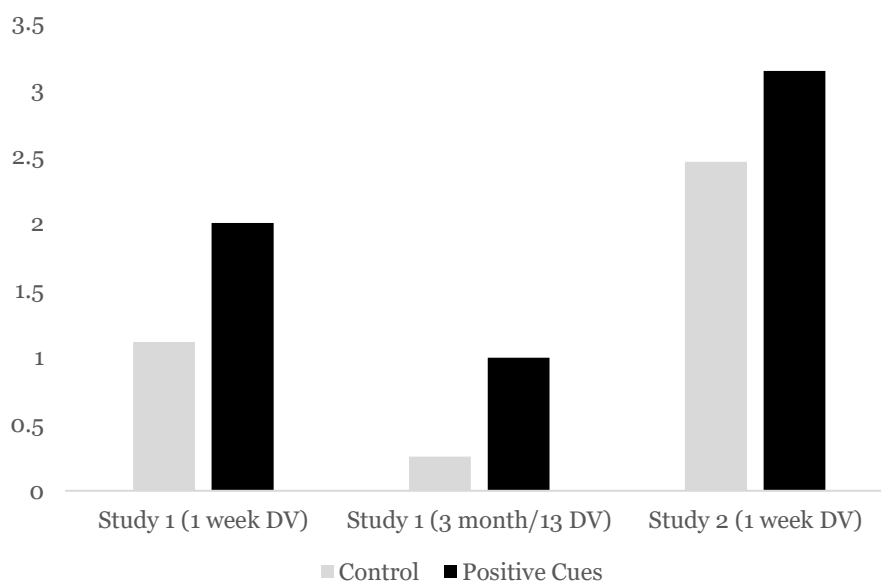
The analysis revealed gender differences in the response to positive cues. We expected that the treatment would be stronger for men, as was indicated with weak significance in Study 1. Furthermore, we hypothesized ( $\gamma$ : social desirability bias will moderate the gender effect of the treatment) which research suggests is more likely to be women (Dalton & Ortegren, 2011). Therefore, we expected to see an interaction between gender and social desirability bias groups. The results support this, in that men consistently increased their reporting regardless of social desirability bias level, while the positive cues treatment was only effective on women with high social desirability bias. This finding suggests that social desirability bias moderates the gender effect of the positive cues treatment. Overall, the study highlights the importance of considering social desirability bias and gender in understanding the effectiveness of interventions aimed at reducing bias in self-reports of undesirable behaviors. Future research could delve deeper into the underlying mechanisms behind these gender differences and explore additional factors that may influence the impact of positive cues in different populations.

## Comparing Study 1 and Study 2

Figure 3 shows a comparison between the positive cues group and the control group in both Study 1 and Study 2, focusing on the mean reported piracy. In Study 1, the figure distinguishes between the two groups based on the time period of the dependent variable question. It is important to note that piracy levels can vary depending on the availability of recent content, such as high-profile movies or albums.

The figure demonstrates a consistent pattern where participants in the positive cues group report higher levels of piracy compared to those in the control group. Notably, the 3-month DV (adjusted to represent a weekly average) indicates the lowest level of average weekly piracy, suggesting the influence of recall bias on reporting. When considering all groups together ( $N = 1,734$ ), the positive cues group reported an average weekly piracy frequency of 1.37, whereas the control group reported 0.89, representing a 42% difference in piracy levels. This highlights the impact of the positive cues treatment in increasing the reported frequency of piracy.

*Figure 3: Mean differences between the piracy reported in the control and the positive cues treatment*



## General Discussion and Theoretical Implications

This paper investigates the relationship between survey prime cues, gender, and social desirability bias in influencing reporting of undesirable behaviors. The findings demonstrate that the positive cues treatment is effective in participants with high social desirability bias, but has no impact on those with low social desirability bias. This suggests that the treatment effectively addresses and mitigates social desirability bias. Individuals with low social desirability bias may be simply not underreporting, which would explain why the treatment does not affect them.

Additionally, the positive cues treatment consistently influences men, increasing their reporting in both the high and low social desirability bias groups. In contrast, it only enhances women's reporting in the high social desirability bias group. This aligns with previous research by Dowsett et al. (2018), indicating that men are more likely than women to employ

rationalizations to justify their undesirable behaviors irrespective of how much they are prone to social desirability. This provides an explanation for why rationale-emphasizing primes have a stronger impact on men.

Overall, the results highlight that primes emphasizing past behaviors are more influential than primes focusing on attitudes. This may explain why previous studies utilizing methods such as forgiving language have not effectively reduced social desirability bias. Furthermore, the persuasive psychology literature suggests that past behavior can be leveraged to shape future behavior, driven by a preference for consistency as demonstrated in the "foot in the door" technique by Freedman and Fraser (1966). Further research is warranted to compare the relative effectiveness of priming attitudes versus behaviors. Nonetheless, this study provides promising evidence regarding the efficacy of behavioral primes in survey contexts.

While we initially theorized that positive cues create cognitive consonance by bridging the gap between moral self-perception and reporting an immoral act, an alternative explanation may lie in the concept of moral licensing. Moral licensing suggests that individuals are driven not by a desire for consistency, but rather by a desire to reach a moral quota, beyond which they feel justified in acting in a self-interested manner (Jordan, Mullen & Murnighan, 2011). Engel and Szech (2020) found that people use box-ticking rationalizations to make unethical choices. When individuals choose a product based on an indicated virtue, such as manufacturing practices, they may be more willing to disregard other ethical considerations, such as working conditions. Sachdeva, Iliev, and Medin (2009) discovered that individuals who narrated positive stories about themselves were less inclined to donate to charity. Monin and Miller (2001) showed that individuals who engaged in or expressed support for moral actions, such as advocating for equal rights, may subsequently feel they have acquired moral "credentials" that allow them to engage in immoral behaviors, such as discriminatory acts towards women or ethnic minorities. Additionally, Mazar and Zhong (2010) conducted an experiment in which consumers were more likely to engage in stealing and lying if they were assigned to buy goods from an online shop with positive credentials compared to a conventional store. This suggests that the moral licensing effect can extend vicariously, where individuals feel justified in acting immorally if someone within their group has already established pro-rights credentials. However, whether the moral licensing/credentials effect works retrospectively, allowing individuals to license their past bad behavior with past good behavior, remains unexplored. If such an effect exists, it could provide an alternative explanation for the observed impact of the positive cues treatment on the reports of undesirable behaviors.

Further research is needed to explore the potential role of moral licensing in the relationship between positive cues and increased reporting of immoral/undesirable behaviors. This would shed light on the underlying mechanisms and provide a more comprehensive understanding of the effects observed in the studies.

## Managerial Implications

Overall, the research underscores the significance of positive cues in increasing honest reporting of undesirable behaviors and sheds light on gender differences in response to survey primes. These findings have practical implications for survey design, communication development, and behavior change strategies in various domains.

The findings from both Study 1 and Study 2 highlight that both men and women underreport undesirable behaviors, the extent of which depends on gender and survey context. This may undermine past research suspected to be subject to social desirability bias, especially if the research explicitly compares men's and women's behavior. This paper finds that framing common behaviors in a positive light influences survey respondents to report negative



behaviors more honestly. This intervention is easily adopted by market researchers and may extend beyond improved past behavior reporting to include attitudes and intentions. Moreover, the method may have broader implications for eliciting truthful responses from individuals involved in more serious criminal or sensitive activities.

These research findings align with existing literature, such as Rothgerber (2013), which suggests that individuals, particularly men, employ rationalizations to overcome cognitive dissonance associated with undesirable behavior. This suggests that marketing communications endeavoring to promote ethical products or services may benefit from identifying the rationales consumers, particularly male ones, use to justify unethical consumption. This would help them design and promote counterarguments

The results indicate that men and women respond differently to survey primes. Men, in particular, tend to be more sensitive to negative primes, resulting in reduced reporting when multiple unethical questions are included in a survey. Marketing survey designers that include sequential undesirable behaviors should be cautious, as our findings suggest that respondents, especially males, may underreport cumulatively. Consequently, it is advisable to tailor survey designs differently for men and women. For example, simply randomizing sensitive survey items may increase survey validity, especially in the case of men.

Lastly, the impact of positive cues on behavior extends beyond improved undesirable behavior reporting and has broad implications. Previous studies, such as Cornelissen et al. (2008) and Hoffmann and Plotkina (2021), have highlighted the potential of positive cues in influencing positive behaviors. This current study demonstrates the power of positive cues in promoting honest reporting. Given the simplicity of implementation (framing common behaviors positively), positive cues can serve as a cost-effective approach for managers seeking to bring about behavioral change.

## Limitations and Future Research

This study is subject to several limitations that should be considered. Firstly, the lack of access to observed undesirable behavior is a significant limitation. As immoral/undesirable behaviors such as piracy are difficult to trace, the study relies on self-reported data, making assumptions that individuals are more likely to underreport than overreport their piracy activities. It is possible that some participants may take pride in their ability to access content for free, considering it a demonstration of technological prowess or resistance to authority. Consequently, the treatments may not have the intended effect. Comparing self-reported data to observed behavior would provide more robust evidence, but such comparisons are challenging due to the clandestine nature of piracy. Future research could explore the relationship between self-reported and observed piracy behavior, as well as examine survey primes' impact on stated behavior and its alignment with observed behavior in domains where behavior can be directly observed.

Another limitation pertains to the measurement of social desirability bias. While the results of this study show promise in linking social desirability bias measured by the Impression Management Scale to the influence of positive cues treatment, it is worth considering the limitations of social desirability bias scales. Lanz, Thielmann, and Gerpott (2022) argue that SDB scales neither measure bias nor traits accurately. Additionally, Lalwani, Lee, Shrum, and Viswanathan (2023) suggest that social desirability bias scales should be complemented with self-deception measures to account for different motivations underlying biased reporting. Future studies could explore alternative measures or approaches to better understand and explain the influence of positive cues treatment and its relationship with social desirability bias.

Lastly, the study highlights the potential of positive cues as a behavior change method, but it is an under-researched area. While this study focused on the impact of positive cues on rationalizing undesirable behaviors, previous research, such as Cornelissen et al. (2008) and Hoffmann and Plotkina (2021), has demonstrated its effectiveness in influencing desirable behaviors. Future research could further explore innovative ways to implement positive cues and examine their impact on various behaviors. Addressing these limitations would enhance the understanding and applicability of the findings in the field of survey research and behavior change strategies.

## References

- Abelson, R. P., Loftus, E. F., & Greenwald, A. G. (1992). Attempts to improve the accuracy of self-reports of voting. *Questions about questions*, 138-53.
- Akers, R. L. (2011). *Social learning and social structure: A general theory of crime and deviance*. Transaction Publishers.
- Al-Rafee, S., & Cronan, T. P. (2006). Digital piracy: Factors that influence attitude toward behavior. *Journal of Business Ethics*, 63, 237-259.
- Baer, A., Trumpeter, N. N., & Weathington, B. L. (2006). Gender differences in memory recall. *Modern Psychological Studies*, 12(1), 3.
- Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, 6(Dec), 157-161.
- Bem, D. J. (1972). Self-perception theory. In *Advances in experimental social psychology* (Vol. 6, pp. 1-62). Academic Press.
- Bernardi, R. A. (2006). Associations between Hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics*, 65, 43-53.
- Beyer, S. (1998). Gender differences in self-perception and negative recall biases. *Sex Roles*, 38(1-2), 103-133.
- Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (1978). Question threat and response bias. *Public Opinion Quarterly*, 42(2), 221-234.
- Brown, R. P., Tamborski, M., Wang, X., Barnes, C. D., Mumford, M. D., Connelly, S., & Devenport, L. D. (2011). Moral credentialing and the rationalization of misconduct. *Ethics & Behavior*, 21(1), 1-12.
- Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, 60(3), 345-375.
- Charles, J. L., & V. Dattalo, P. (2018). Minimizing social desirability bias in measuring sensitive topics: The use of forgiving language in item development. *Journal of Social Service Research*, 44(4), 587-599.
- Cheng, H. K., Sims, R. R., & Teegen, H. (1997). To purchase or to pirate software: An empirical study. *Journal of Management Information Systems*, 13(4), 49-60.
- Christensen, A. L., & Eining, M. M. (1991). Factors influencing software piracy: Implications for accountants. *Journal of Information Systems*, 5(1), 67-80.
- Chung, J., & Monroe, G. S. (2003). Exploring social desirability bias. *Journal of Business Ethics*, 44, 291-302.
- Cornelissen, G., Pandelaere, M., Warlop, L., & Dewitte, S. (2008). Positive cueing: Promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing*, 25(1), 46-55.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349.
- Dalton, D., & Ortegren, M. (2011). Gender differences in ethics research: The importance of controlling for the social desirability response bias. *Journal of Business Ethics*, 103, 73-93.

- Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addictive Behaviors*, 35(4), 302-311.
- Dowsett, E., Semmler, C., Bray, H., Ankeny, R. A., & Chur-Hansen, A. (2018). Neutralizing the meat paradox: Cognitive dissonance, gender, and eating animals. *Appetite*, 123, 280-288.
- Duff, B., Hanmer, M. J., Park, W. H., & White, I. K. (2007). Good excuses: Understanding who votes with an improved turnout question. *Public Opinion Quarterly*, 71(1), 67-90. Kalton, Collins, and Brook (1978)
- Engel, J., & Szech, N. (2020). A little good is good enough: Ethical consumption, cheap excuses, and moral self-licensing. *PloS one*, 15(1), e0227036.
- Festinger, L. (1957). Social comparison theory. *Selective Exposure Theory*, 16, 401.
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and social psychology*, 4(2), 195.
- Fukukawa, K., Zaharie, M. M., & Romonți-Maniu, A. I. (2019). Neutralization techniques as a moderating mechanism: Ethically questionable behavior in the Romanian consumer context. *Psychology & Marketing*, 36(2), 138-149.
- Gendall, P., Hoek, J., & Blakeley, M. (1992). Estimating a socially undesirable behaviour. *Marketing Bulletin*, 3, 1-8.
- Gergely, M., & Rao, V. S. C. (2014). Social Desirability Bias in Software Piracy Research. In *ECIS*.
- Glover, S. H., Bumpus, M. A., Sharp, G. F., & Munchus, G. A. (2002). Gender differences in ethical decision making. *Women in Management Review*, 17(5), 217-227.
- Harrison, D. A., & McLaughlin, M. E. (1993). Cognitive processes in self-report responses: tests of item context effects in work attitude measures. *Journal of Applied Psychology*, 78(1), 129.
- Higgins, G. E. (2006). Gender differences in software piracy: The mediating roles of self-control theory and social learning theory. *Journal of Economic Crime Management*, 4(1), 1-30.
- Hoffmann, A. O., & Plotkina, D. (2021). Positive framing when assessing the personal resources to manage one's finances increases consumers' retirement self-efficacy and improves retirement goal clarity. *Psychology & Marketing*, 38(12), 2286-2304.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37-67.
- 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.
- Kim, S. H., & Kim, S. (2016). Social desirability bias in measuring public service motivation. *International Public Management Journal*, 19(3), 293-319.
- King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, 17(2), 79-103.
- Kingston, N. M., & Dorans, N. J. (1984). Item location effects and their implications for IRT equating and adaptive testing. *Applied Psychological Measurement*, 8(2), 147-154.
- Klar, S., Weber, C. R., & Krupnikov, Y. (2016, December). Social desirability bias in the 2016 presidential election. In *The Forum* (Vol. 14, No. 4, pp. 433-443). De Gruyter.
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201-219.

Lalwani, A. K., Lee, H., Shrum, L. J., & Viswanathan, M. (2023). Men engage in self-deceptive enhancement, whereas women engage in impression management. *Psychology & Marketing*, 40, 1405–1416.

Light Space Time (n.d.) *10 Ways that We Can Support the Arts* Retrieved May 26, 2023, from [lightspacetime.art/10-ways-that-we-can-support-the-arts-and-artists](https://lightspacetime.art/10-ways-that-we-can-support-the-arts-and-artists)

Luchs, M. G., & Mooradian, T. A. (2012). Sex, personality, and sustainable consumer behaviour: Elucidating the gender effect. *Journal of Consumer Policy*, 35, 127-144.

Lanz, L., Thielmann, I., & Gerpott, F. H. (2022). Are social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior. *Journal of Personality*, 90(2), 203-221.

Larson, R. B., & Kinsey, J. (2019). Culture and sampling issues with “Green” attitude research. *Social Marketing Quarterly*, 25(2), 91-106.

Mazar, N., & Zhong, C. B. (2010). Do green products make us better people? *Psychological Science*, 21(4), 494-498.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, 65(2), 131-149.

Morris, R. G., Johnson, M. C., & Higgins, G. E. (2009). The role of gender in predicting the willingness to engage in digital piracy among college students. *Criminal Justice Studies*, 22(4), 393-404.

Näher, A. F., & Krumpal, I. (2012). Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. *Quality & Quantity*, 46, 1601-1616.

Paulhus, D.L. (1991) Measurement and control of response bias. *Measures of Personality and Social Psychological Attitudes*, 1, 17–59

Phau, I., Teah, M., & Liang, J. (2016). Investigating the factors influencing digital movie piracy. *Journal of Promotion Management*, 22(5), 637-664.

Piazza, J., Ruby, M. B., Loughnan, S., Luong, M., Kulik, J., Watkins, H. M., & Seigerman, M. (2015). Rationalizing meat consumption. The 4Ns. *Appetite*, 91, 114-128.

Prolific (2023) *Conduct your best research* [www.prolific.co/researchers](http://www.prolific.co/researchers) ([www.prolific.co](http://www.prolific.co))

Riekkinen, J. (2016). *Dissonance and Neutralization of Subscription Streaming Era Digital Music Piracy: An Initial Exploration*. In PACIS 2016: Proceedings of the 20th Pacific Asia Conference on Information Systems (Article 251). Association for Information Systems.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, 2(4), 320-332.

Rothgerber, H. (2013). Real men don't eat (vegetable) quiche: Masculinity and the justification of meat consumption. *Psychology of Men & Masculinity*, 14(4), 363.

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.

Schwarz, N., & Bles, H. (1992). Scandals and the public's trust in politicians: Assimilation and contrast effects. *Personality and Social Psychology Bulletin*, 18(5), 574-579.

- Sintov, N. D., & Prescott, C. A. (2011). The influence of social desirability and item priming effects on reports of proenvironmental behavior. *Ecopsychology*, 3(4), 257-267.
- Smallridge, J. L., & Roberts, J. R. (2013). Crime Specific Neutralizations: An Empirical Examination of Four Types of Digital Piracy. *International Journal of Cyber Criminology*, 7(2).
- Soo, N. (2020, April 3) *8 Ways You Can Support Your Local Arts The Artling*  
[theartling.com/en/artzine/support-local-arts-community](http://theartling.com/en/artzine/support-local-arts-community)
- Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, 37(3), 228-243.
- Steele, C. M., & Liu, T. J. (1981). Making the dissonant act unreflective of self: Dissonance avoidance and the expectancy of a value-affirming response. *Personality and Social Psychology Bulletin*, 7(3), 393-397.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. Jossey-Bass,
- Swann Jr, W. B., & Schroeder, D. G. (1995). The search for beauty and truth: A framework for understanding reactions to evaluations. *Personality and Social Psychology Bulletin*, 21(12), 1307-1318.
- Tajfel, H. and Turner, J.C. (1985) 'The Social Identity Theory of Group Behavior'. In Tajfel, H.(ed.)*Psychology of Intergroup Relations*. Cambridge: Cambridge University Press, pp. 15-40.
- Taylor, S. A. (2012). Evaluating digital piracy intentions on behaviors. *Journal of Services Marketing*, 26(7), 472-483.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2), 275-304.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Vinten, G. (1995). Open versus closed questions—an open issue. *Management Decision*, 33(4), 27-31.

## Appendix A: Common Behavior Pilot

A common behavior pilot (N = 100) was conducted to decipher which common positive entertainment consumption behaviors participants; *i*) agreed they participated in, *ii*) agreed were supportive of the arts, and *iii*) reported that they made them feel ethically positive.

Nine positive behaviors were identified from researching suggestions of ways the public can support the creative industries on various creative industry/trade body websites (such as [lightspacetime.art](http://lightspacetime.art) and [getitrightfromagenuinesite.org](http://getitrightfromagenuinesite.org)). Participants were asked whether they strongly agreed (1) to strongly disagreed (5) that they participated in a behavior (for example subscribing to a legal streaming service) “common behaviors” and whether they believed the behavior supported the arts “support”. They were then asked to report on a 5-point Likert scale whether a statement that combined the two, such as “I support the arts by subscribing to a TV/music entertainment service/s (such as Netflix or Spotify)” made them feel ethically positive (1) or negative (5) “positive behaviors”.

For the statement to make it into the main study, the statements had to meet the criteria of averaging between 1 (strongly agree) and 3 (neither agree nor disagree) for the “common behaviors” and “support” questions. And they had to average between 1 (positive) and 3 (neither positive nor negative) for the “positive behaviors”. This left five positive behaviors which participants broadly agreed they did, were supportive to the creative industries, and made them feel ethically positive (see *table A1*).

Table A1: Behaviors tested in the “common behaviors” pilot

Statement	Mean agreed participants performed the behavior	Mean agreed the behaviors supported the arts	Mean agreed the statements made them feel ethically positive	Result
I support the arts by going to the cinema, concerts, or theatre	2.03	1.68	1.92	Met the Criteria
I support the arts by visiting museums, cultural heritage sites or galleries	2.16	1.42	1.75	Met the Criteria
I support the arts by encouraging friends/family members to go to local art/theatre/exhibition/music events	2.72	1.63	2.14	Met the Criteria
I support the arts by subscribing to TV/music entertainment service/s (such as Netflix or Spotify)	1.72	2.78	2.44	Met the Criteria
I support the arts by sharing music/film/TV releases I like on my social media page/s	3.83	2.18	2.18	Did not meet the criteria
I support the arts by writing positive reviews to show support for an artist's work/event on sites like Etsy, Facebook or Google	3.03	1.66	2.23	Did not meet the criteria
I support the arts by purchasing physical products from artists/musicians such as records, merchandise or pictures	3.14	1.50	2.04	Did not meet the criteria
I support the arts by buying artwork directly from the artist from time to time (such as from craft fairs, local shops or websites like Etsy) rather than always using large retailers such as Amazon	2.76	1.42	2.13	Met the Criteria
I support the arts by donating money to a gallery/theatre/music venue or charity	3.11	1.59	2.28	Did not meet the criteria



The final five behaviors which met the criteria were similar to the ones used in Study 1, as can be seen in *Table A2*.

*Table A2: Behaviors tested in the “common behaviors” pilot*

Study 1 Positive Cues Questions	Study 2 Positive Cues Questions
<ol style="list-style-type: none"> <li>1. I support the arts by purchasing tickets for the cinema or live events</li> <li>2. I support the arts by purchasing legal TV/music subscriptions (such as Netflix or Spotify)</li> <li>3. I support creative talent by sharing an artist’s film/music release on my social media page</li> <li>4. I help promote creative content by recommending a film/music release to people I know</li> </ol>	<ol style="list-style-type: none"> <li>1. I support the arts by visiting the cinema, concerts or theatre</li> <li>2. I support the arts by subscribing to TV/music entertainment service/s (such as Netflix or Spotify)</li> <li>3. I support the arts by visiting museums, cultural heritage sites or galleries</li> <li>4. I support the arts by encouraging friends/family members to go to local art/theatre/exhibition/music events</li> <li>5. I support the arts by buying artwork directly from the artist from time to time (such as from craft fairs, local shops or websites like Etsy) rather than always using large retailers such as Amazon</li> </ol>

## Appendix B - Study 1 Additional Analyses

Table B1: Treatment groups vs the neutral condition, linear regression

Models	1	2	3	4	5	6
				Male	Female	All
Male	0.564*** (0.164)	0.558*** (0.164)	1.032** (0.460)			
Permissive		-0.354 (0.251)	-0.164 (0.119)	-0.580 (0.502)	-0.164 (0.119)	-0.317 (0.251)
Restrictive		-0.321 (0.261)	0.314 (0.275)	-0.984** (0.462)	0.314 (0.275)	-0.296 (0.258)
Positive Cues		0.730* (0.426)	0.612 (0.499)	0.797 (0.703)	0.612 (0.499)	0.744* (0.424)
Contrast		-0.401* (0.232)	0.0343 (0.154)	-0.893* (0.462)	0.0343 (0.154)	-0.389* (0.232)
Control		-0.0870 (0.263)	0.0799 (0.160)	-0.289 (0.518)	0.0799 (0.160)	-0.0597 (0.262)
Permissive x Male			-0.417 (0.516)			
Restrictive x Male			-1.298** (0.538)			
Positive Cues x Male			0.185 (0.862)			
Contrast x Male			-0.927* (0.487)			
Control x Male			-0.369 (0.542)			
Constant	0.417*** (0.0946)	0.495** (0.199)	0.275** (0.112)	1.307*** (0.447)	0.275** (0.112)	0.751*** (0.218)
Observations	1,156	1,156	1,156	577	579	1,161
Adjusted R-squared	0.009	0.024	0.028	0.027	0.003	0.014

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table B3: Linear regression analysis comparing piracy reported in 3months (13 weeks) to 1-week dependent variables

Model	1	2	3	4 Male	5 Female
Pilot One	0.479*** (0.164)	0.480*** (0.162)	0.857*** (0.282)	1.218** (0.491)	0.440* (0.225)
Permissive		-0.252 (0.190)	0.0534 (0.130)	0.148 (0.240)	-0.0638 (0.0766)
Restrictive		-0.230 (0.199)	-0.0782 (0.118)	-0.285** (0.144)	0.149 (0.210)
Positive Cues		0.810** (0.390)	0.741** (0.368)	1.463** (0.704)	-0.00288 (0.106)
Contrast		-0.323** (0.165)	0.103 (0.131)	0.0655 (0.226)	0.153 (0.144)
Neutral		0.0723 (0.260)	0.401 (0.366)	0.838 (0.782)	0.0643 (0.166)
Permissive x Pilot One			-0.603 (0.376)	-0.822 (0.672)	-0.358 (0.240)
Restrictive x Pilot One			-0.296 (0.395)	-0.692 (0.559)	0.107 (0.518)
Positive Cues x Pilot One			0.150 (0.780)	-0.692 (1.196)	1.121 (1.020)
Contrast x Pilot One			-0.846*** (0.323)	-1.274** (0.547)	-0.394 (0.312)
Neutral x Pilot One			-0.659 (0.515)	-1.037 (1.023)	-0.277 (0.318)
Constant	0.459*** (0.0910)	0.447*** (0.147)	0.254*** (0.0767)	0.378*** (0.141)	0.135** (0.0603)
Observations	1,161	1,161	1,161	577	579
Adjusted R-squared	0.006	0.021	0.021	0.026	0.016

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## Appendix C - Study 2 Additional Analyses

Figure C2: Margins plot from linear regression comparing low vs high social desirability bias (SDB) groups and the effect of the positive cues treatment

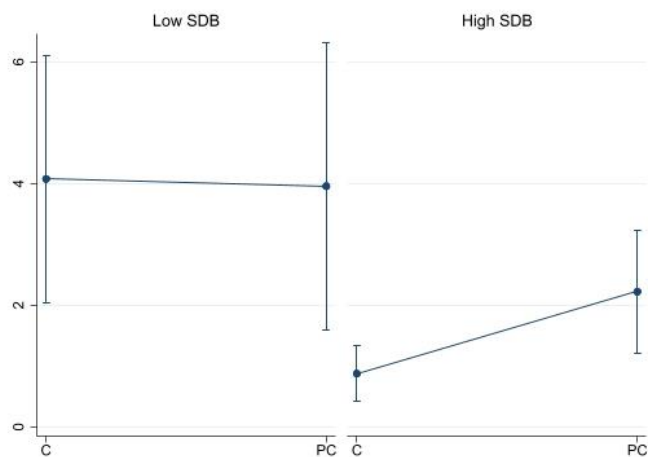


Figure C3: Margins plot from linear regression comparing social desirability bias (SDB), gender, and treatment groups

