CONSUMER ENGAGEMENT IN THE DIGITAL ERA

by

CHI PHUONG TRAN

A DISSERTATION

Presented to the Department of Marketing,

and the Division of Graduate Studies of the University of Oregon

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Chi Phuong Tran

Title: Consumer Engagement in the Digital Era

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Marketing by:

Joshua T. Beck	Co-chair
Hong Yuan	Co-chair
Conor Henderson	Core Member
Jocelyn Hollander	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Chi Phuong Tran

Doctor of Philosophy

Department of Marketing, Lundquist College of Business

June 2023

Title: Consumer Engagement in the Digital Era

The past few decades have witnessed massive expansion of technology infrastructure and digital devices. As technological advancement continues to upend many existing social and business norms, it fundamentally changes how we as consumers participate in the marketplace: from the way we communicate and interact with each other and with companies to the way we are exposed to different identities while continuously developing and maintaining our own in the digital age. In this dissertation, I investigate how consumers view and react to minority identities and how we strive to protect our own personal information as we become active members of the digital world. Building on foundational theories in sociology, consumer psychology and marketing strategy and utilizing a multi-method approach (econometrics models, text analysis and experiments), this two-essay dissertation strives to provide insights for market stakeholders (consumers, marketers, policymakers) who are navigating a digitally advanced, increasingly diverse and complex demographic and sociopolitical environment.

Essay I ("Shifting Frames: How #MeToo Shaped Consumers' Movie-Going Decisions"; under review at the *Journal of Marketing*) capitalizes on a rich data set of top 1500 performing movies from 2010 to 2019 and investigates how consumers perceive and respond to

marginalized identities in an increasingly diverse and digitally driven world. Specifically, this first essay investigates how major social justice movements that primarily aim to change the narratives around marginalized identities could spill over to affect many other consumer spending decisions. Essay II ("Feeling Violated: How and When Privacy Violations Produce Commensurate Consumer Responses;" in preparation for 3rd round revision at the *Journal of Consumer Research*) studies how consumers strive to protect their personal identity and information in the pervasive digital environment.

ACKNOWLEDGMENTS

I would like to address my deepest gratitude to Professors Joshua T. Beck and Hong Yuan for their guidance, support and mentorship throughout my Ph.D. journey. Their expertise, feedback and encouragement were instrumental in guiding the development of this thesis.

I am also grateful to the members of my thesis committee, Professors Conor Henderson, John Clithero and Jocelyn Hollander for their insightful comments, suggestions and feedback that help improve this work.

I extend my heartfelt thanks to Professors Brandon J. Reich at Portland State University and Nooshin L. Warren at the University of Arizona, whose assistance, guidance, encouragement and fruitful discussions enriched my research work.

This research was financially supported in part by the Oregon Consumer Protection Grant; the John and Emiko Kageyama Endowment Fund; and the Roger Best Research Funds.

Finally, I am deeply indebted to my family, especially my beloved husband, and my friends for their unwavering love, support, and encouragement. Their understanding and motivation helped me to overcome the challenges of pursuing a Ph.D.

For my husband who has an unlimited capacity for listening to my research ideas and (mostly) frustrations and challenges.

For my mother who taught me to find my own voice and advocate for what I believe in, no matter how difficult the situation.

For my father whose only way out of poverty was academia, whose love for knowledge and learning continues to motivate me every day.

Lastly, for my advisor, Josh, whose wisdom and intellect have challenged me to think deeply and critically, whose passion for the field and for research has inspired me to find my own, whose advice, dedication and mentorship I will always miss dearly. It is truly a great privilege to be your student, Josh.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESSAY I:
SHIFTING FRAMES: HOW #METOO SHAPED CONSUMERS' MOVIE-GOING DECISIONS

In recent years, social justice movements such as the Occupy Wall Street Movement, Black Lives Matter, and #MeToo have dominated the public stage. Catalyzed by advances in technology and media platforms, social movements have become increasingly prevalent in consumers' daily experiences (Amenta and Polletta 2019; Della Porta and Diani 2020). For example, in the month following George Floyd's death, Black Lives Matter (BLM) was tweeted 47.8 million times (Anderson et al. 2020), and as many as 26 million people participated in BLM rallies (Buchanan, Bui, and Patel 2020). In contrast to boycotts and buycotts, which have been extensively studied (Klein, Smith, and John 2004; Kozinets and Handelman 2004; McDonnell and King 2013; Sen, Gürhan-Canli, and Morwitz 2001), marketing scholars are only recently considering how social justice movements may alter consumer spending decisions (Nardini et al. 2021; Della Porta and Diani 2020). In the present work, we contribute to this emerging discourse by examining how #MeToo influenced the box office performance of US films.

Historically, many film production studios have created films with characters that conform to common gender stereotypes (e.g., heroic male, damsel in distress). This is because such stereotypes are thought to reflect audience preferences (Grau and Zotos 2016). The #MeToo movement, which was centered on greater female empowerment and male accountability related to sexual misconduct and other abuses of power, did not expressly aim to change consumers' movie preferences. However, through a collaborative, distributed, and iterative process, #MeToo formulated a clear collective action frame that highlighted and identified problems (diagnostic framing), proposed solutions (prognostic framing), and provided additional rationale and personal appeals (motivational framing). We posit that as this collective action frame diffused and prevailed in public opinion, "winning hearts and minds" (Nardini et al. 2021, p. 115), it had the effect of shifting how movie-goers evaluated gender stereotypes in films. As a result, in the immediate aftermath of #MeToo, films with varying gender portrayals performed better or worse, depending on how those stereotypes or counter-stereotypes aligned with #MeToo's collective action frame.

To examine this, we used computer programing to collect data from IMDb, Box Office Mojo, and The-Numbers.com and hand-coded variables to assemble a unique and comprehensive dataset of the top-performing 1,500 films from 2010 to 2019 (150 films annually) in the US. Excluding documentaries, animations, and non-English speaking films, our final dataset consists of details regarding 1,326 films and 3,494 main actors, directors, and writers for these films. For our empirical investigation, we treat #MeToo as an exogenous shock and analyze box office performance of female (vs. male)-centric films prior to and after #MeToo using year and studio fixed-effects models with a wide range of controls to account for endogeneity concerns. We utilize topic modeling on 54,725 unique crowdsourced movie keywords to uncover existing gender (counter-) stereotypes portrayed in films. This creative approach using keywords (as opposed to scripts or summaries) reduces computational difficulties often associated with natural language processing procedures, while ensuring the themes uncovered are reflective of the audience's perspective, enhancing ecological validity. We further supplement our analyses with propensity score matching to address sample imbalance, and we provide a comprehensive set of sensitivity and robustness checks.

Through our rigorous empirical approach we find that, consistent with the logic of the movement, #MeToo decreased the performance of films containing negative female stereotypes (weak, victimized, emotional) by approximately $8 million for an average female-led film and enhanced the performance of films containing positive male counter-stereotypes (warm, gentle, non-powerful or violent) by approximately $12 million for an average male-led film. Together, our findings shed light on how social movements such as #MeToo unfold in the public and immediately affect consumers' spending decisions.

Our research extends work on social movements, consumer decision-making, and marketing strategy in several ways. First, by studying the #MeToo movement and the film industry, we demonstrate how disruptive social movements shape culture through framing and frame diffusion, thereby undermining the attractiveness of traditional stereotypes. In doing so, we contribute to the literature on stereotype correction and adjustment (Ellemers 2018; Hilton and Von Hippel 1996; Judd and Park 1993). Second, our research is the first to empirically examine and establish the spillover impacts of large social justice movements on consumer behavior, broadening the understanding of social movements in the consumer domain outside of boycotts or buycotts. Finally, bridging research on social movements in sociology with the consumer psychology literature, we provide a conceptual framework through which the impacts of other social movements on consumers could be analyzed and understood at aggregate societal level. As social movements continue to proliferate and thrive, our findings provide valuable managerial insight, particularly for entertainment industry stakeholders who are determining casting and content decisions in an ever-changing sociopolitical environment.

Conceptual Background

Social Movements and Collective Action Frames

Social movements are the product of informal, networked interactions among individuals, groups, and organizations that enact societal change (DeFronzo and Gill 2019; Diani 1992; Turner 1969). They are disruptive in ways that may involve broad structural change (e.g., the Arab Spring) or narrower behavioral affordances (e.g., marijuana legalization). In recent years, notable social movements have sought to create new empowerment narratives for marginalized groups by challenging social norms and generating awareness and understanding (Benford and Snow 2000; DeFronzo and Gill 2019). Originally considered spontaneous and ephemeral, social movements have become more persistent and better organized by distributed media tools (e.g., mobile phones and social media) that efficiently mobilize resources and support across a large number of people (McCarthy and Zald 1977; Morris 2000; Della Porta and Diani 2020).

Critically, all social movements develop collective action frames, which are the "schemata of interpretation" that supply the logic of the social movement and permit newcomer participation (Benford and Snow 2000; p.614). Similar to the notion of a narrative (Nardini et al. 2021), these action frames tactically shape meanings of prominent events, evoke emotions, and inspire actions (Benford and Snow 2000; Jasper 2011; Della Porta and Diani 2020). Collective action frames serve three primary functions: highlighting and identifying problems (diagnostic framing), proposing solutions (prognostic framing), and providing additional rationale and personal appeal (motivational framing). Most research in collective action frames focuses on their strategic deployment from social movement organizations and activists to the public (Benford and Snow 2000). However, less is known about frames' diffusion process in which the logic and ideas of social movements spread, evolve, and transform among the public (Benford and Snow 2000; Reinecke and Ansari 2021). The effect of social movements and their collective action frames on consumer spending decisions, therefore, remains an intriguing empirical question, despite much research on social movements' impact on government policy and corporate practices (Amenta and Polletta 2019; Della Porta and Diani 2020).

Collective Action Frames and Consumer Frame Shift

As social movements and their collective action frames are increasingly amplified and widely distributed by news and social media (Amenta and Polletta 2019), it is critical to understand how social movements impact consumers. Past work has examined boycotts and buycotts as specific forms of social movements intended to reinforce or change particular corporate practices—e.g., companies' practices related to reproductive rights, racial discrimination, and labor (Klein, Smith, and John 2004; Kozinets and Handelman 2004; McDonnell and King 2013; Sen, Gürhan-Canli, and Morwitz 2001). Here, activists attempt to persuade consumers to reward or punish companies based on their practices (Klein, Smith, and John 2004). What about the consumer effects of social movements that are not expressly consumption oriented? For example, the #MeToo, Black Lives Matter, and Gay Pride movements sought governance and policy changes that were outside of consumer domains. However, recent work posits that consumer psychology may be impacted by social movements via pathways that differ from traditional boycotting (Nardini et al. 2021).

In this research, we posit that social movements, through their collective action frames, create a consumer frame shift that is aligned with the social movement's goals. Indeed, as Nardini et al. (2021) formalize, social movements can have the effect of winning consumers' hearts and minds. A litany of research establishes framing effects (Janiszewski and Wyer Jr 2014; Smith and Petty 1996; Tversky and Kahneman 1981). Frames shape consumer reference points by making certain attributes more or less noticeable, which subsequently affects consumer evaluation and purchase. In social movements, the collective action frames are communicated and made salient by widespread news coverage and viral appearance on social media (Amenta and Polletta 2019). These collective action frames may change even stubborn attitudes. For example, the Black Lives Matter movement publicized new perspectives and the consideration of personal bias, which past work has shown can alter social perceptions (Hilton and Von Hippel 1996; Judd and Park 1993). While social movements' primary goals are to mobilize resources and inspire public actions (e.g., protest), social movements also precipitate cultural production (Diani 1992). Comparably, we argue for a frame diffusion process, whereby the logics of the collective action frame are distributed to consumers and subsequently alter their worldviews in a way that is consistent with the collective action frame. We suggest, then, that social movements can alter consumer behaviors even when a movement has no such express aim. We examine these theoretical propositions in the context of #MeToo and the US film industry.

Context: #MeToo and the US Film Industry

The #MeToo movement is one of the most prominent social movements in the US (and in the world) in recent years (Xiong, Cho, and Boatwright 2019). While #MeToo was initiated by activist Tarana Burke in 2006, the movement gained national attention in October of 2017. On October 5th, explosive reports of sexual violence committed by then prominent Hollywood producer Harvey Weinstein made national news. On October 15, 2017, actress Alyssa Milano tweeted to her three million followers to encourage survivors of sexual violence to include the hashtag #MeToo in their replies. As many other celebrities and millions of people soon joined Milano in their replies, #MeToo quickly became viral in the US and around the world. From a hashtag, #MeToo gained significant momentum to become a social movement against sexual violence, highlighting the prevalence and consequences of gender injustice. #MeToo was used 19 million times on Twitter in the year following the movement's takeoff. Two out of three American social media users reported having seen or participated in the #MeToo discussion on their feed (Pew Research Center 2018). In the same year, TIME magazine named #MeToo activists, the Silence Breakers, as the TIME's Person of the Year.

While most research on #MeToo has focused on its virality and dominance in the news (Pew Research Center 2018) and its effects in the workplace or in behavior reporting (Atwater et al. 2021), #MeToo also provides a great context for the study of social movement's influence on consumers. First, #MeToo's public rise is undoubtedly deeply rooted in the motion picture (entertainment) industry (Luo and Zhang 2021). The movement was ignited and amplified by Hollywood actresses who shared their experiences working in films (Xiong, Cho, and Boatwright 2019). Second, the social problems highlighted by #MeToo were intrinsically relevant to the existing issues within filmmaking organizations (Erigha 2015; Grau and Zotos 2016). Third, the motion picture industry plays a critical role in reflecting and shaping cultural trends (Eliashberg, Elberse, and Leenders 2006). As #MeToo's collective action frame diffused to the public, its ideas and values transformed people's daily interactions and decisions (Amenta and Polletta 2019). We posit that this frame shift affected consumer preferences.

Hypothesis Development

#MeToo's Collective Action Frames and Precipitating Conditions in Films

The #MeToo social movement and its associated messaging represent one of the most powerful collective action frames in the past few decades (McDonald 2019; Xiong, Cho, and Boatwright 2019). #MeToo has three core elements: one, raising awareness about the overwhelming prevalence of sexual violence, especially in the workplace; two, empowering survivors by providing support and breaking gender stereotypes and victim stigma; and three, advocating for systematic changes towards sexual violence ("MeToo Movement" 2021). These three frame elements respectively represent the diagnostic, prognostic, and motivational components of the #MeToo collective action frame (Benford and Snow 2000). The massive volume of #MeToo-related tweets, news and media segments, as well as daily conversations further solidify and legitimize these frames (Xiong, Cho, and Boatwright 2019). #MeToo creates a positive narrative that destigmatizes survivors of sexual violence, the majority of whom are women, depicting these women as brave, empowered, and heroic silence breakers who were once “stoic victims of an unjust system” (Starkey et al. 2019, p. 437). These depictions are counter to the persistent gender stereotypes that are often reflected in and reinforced by films (Grau and Zotos 2016; Xu et al. 2019).

Prior to #MeToo, the film industry was known for its lack of gender equality (Erigha 2015). There existed a significant wage gap between male and female actors, female directors and writers had significantly less representation and opportunity, and most industry executives were male (Luo and Zhang 2021). The significant power-gender imbalance in Hollywood and the entertainment industry also translated into films that perpetuated gender stereotypes (Ellemers 2018; Grau and Zotos 2016; Xu et al. 2019). These stereotypes persist, industry stakeholders argue, because they reflect dominant social values and are thus profitable (Grau and Zotos 2016). Changes in stereotypical depictions are often unsuccessful because challenging gender stereotypes had been seen as controversial, risky, and costly (Ellemers 2018; Grau and Zotos 2016; Luo and Zhang 2021).

The Effects of #MeToo on Consumer Film Preferences

The #MeToo movement criticizes prevailing gender stereotypes. Research on stereotypes suggests that stereotypes are resistant to change because the process of disconfirming stereotypes requires cognitive effort, can result in disapproval by others, and elevates stress (Ellemers 2018). For individuals to question stereotypes, persuasive disconfirming information about a social group needs to emerge, dramatically or gradually (Ellemers 2018; Hilton and Von Hippel 1996). For the stereotype-perpetuating Hollywood industry, #MeToo and its collective action frame, therefore, was a disruptive shock that challenged the status quo (Carlsen et al. 2018; Luo and Zhang 2021).

Because of #MeToo's strong positive framing of women (McDonald 2019; Starkey et al. 2019), immediately following the rise of #MeToo in 2017, we predict that paradoxically, #MeToo might negatively impact the performance of many female-centric films. This seemingly paradoxical effect occurs because, as compared to #MeToo's positive frames, films prior to #MeToo often portrayed women using negative stereotypes. Many films in Hollywood centralize around (heterosexual) men's narratives (Erigha 2015). In most films, the male is portrayed as agentic, independent, and highly competent living a life full of adventure (Xu et al. 2019). Physically masculine and powerful but often emotionally unavailable, male characters are also more likely to be heroic, professional, and authoritarian (Grau and Zotos 2016; Xu et al. 2019). Negative male characters are often used to highlight positive ones in the same films, and men in counter-stereotypical roles are minimally present on screen (Erigha 2015). By contrast, even in female-centric films (i.e., those with female leads or a high proportion of females in the main cast), the female characters are often depicted as dependent on men and lacking in agency and competence (Miller, Rauch, and Kaplan 2016; Xu et al. 2019). Women are also more likely to be objectified and portrayed as victims in need of male protection (Erigha 2015; Linz, Donnerstein, and Penrod 1988). Even when women have agency (e.g., career woman), they are often portrayed in a negative light: overly aggressive, revengeful, and hostile, only rewarded as they regain their feminine traits of warmth and communality (Ezzedeen 2015; Valdivia 1998). For example, in *Lucy*, a commercially successful 2014 film, Scarlett Johansson portrays the victimized female lead who turns highly violent. We predict that #MeToo has a negative effect on female-centric films that perpetuate negative stereotypes about women. Table 1.1 summarizes positive and negative gender stereotypes in films (Eagly and Mladinic 1989; Ellemers 2018; Williams and Best 1990).

TABLE 1. 1. Positive and Negative Gender (Counter) Stereotypes

	Positive	**Negative**
Stereotypical male traits *(counter-stereotypical female traits)*	Powerful, heroic, competent	Self-focused, violent, cold
Stereotypical female traits (*counter-stereotypical male traits)*	Warm, gentle, other-focused	Weak, victim, emotional

We further anticipate that the #MeToo social movement affects male portrayals. Diagnostically, #MeToo was partially a reaction to violence by powerful men. The collective action frame highlights the extent and severity of the issue, attributing it to the existing culture of gender power asymmetry and victim blaming. Consequently, under this action frame, we expect consumers to shift away from films that negatively portray men as overwhelmingly violent. These portrayals are widely common in Hollywood films prior to #MeToo. For example, the popular three-film series *Taken* features Liam Neeson in a masculine, skilled, but highly violent agent role protecting his helpless wife and daughter from being kidnapped. Notably, the content of male and female stereotypes differs (see Table 1.1). Also, as above, we expect that these effects will be conditioned on whether males are more centrally portrayed in the film. Formally stated:

> **H_1: There is a negative effect of #MeToo on the performance of female-centric films that portray women using negative female stereotypes.**

> **H_2: There is a negative effect of #MeToo on the performance of male-centric films that portray men using negative male stereotypes.**

Additionally, #MeToo's prognostic collective action frame proposes an approach of reframing victims as survivors and advocates for social support and understanding. Under this aspect, women who speak up about their experiences are brave and heroic silence-breakers (Starkey et al. 2019), and men are encouraged to become allies, engaging in active listening and dismantling the existing toxic culture (Flood 2019; PettyJohn et al. 2019). As a result, consumers might be more likely to endorse films that portray women in positive and heroic roles that are counter-stereotypical (e.g., *Wonder Woman*), further opposing the victimized damsel-in-distress narrative. We expect that in encouraging men to become more emotionally understanding, this framing increases consumers' preference for counter-stereotypical and more emotionally complex portrayals of men (e.g., Bradley Cooper's character Jack in *A Star is Born*).

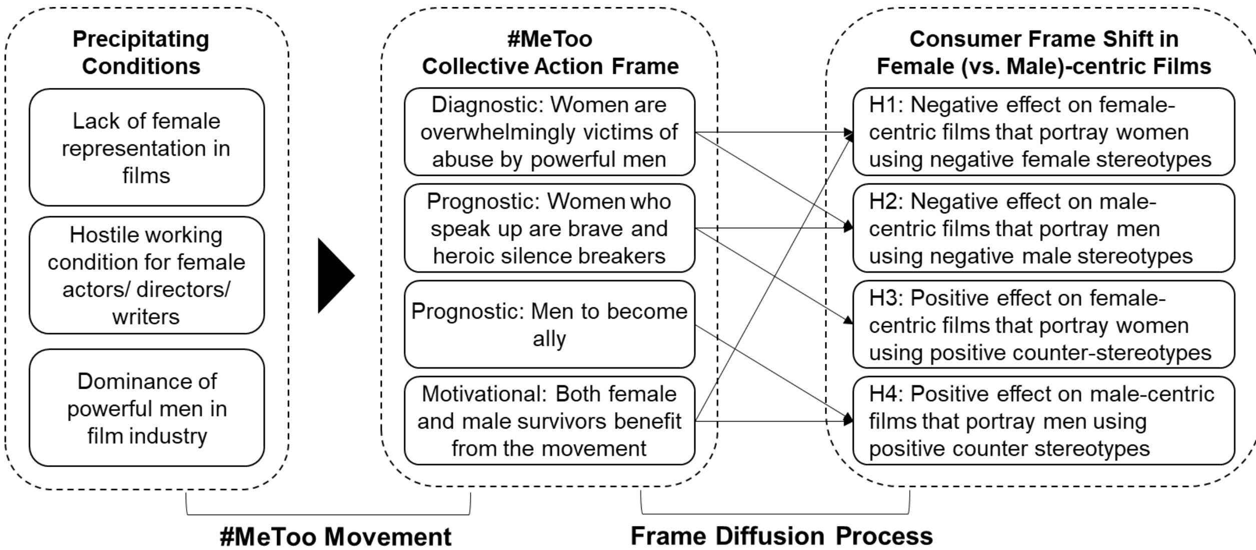
Also, reinforcing both diagnostic and prognostic frames, #MeToo's motivational collective action frames focus on envisioning systematic change in sexual violence, instilling a sense of urgency and efficacy of the movement, scaling up social and moral support (Benford 1993; Xiong, Cho, and Boatwright 2019). Under this framing, #MeToo's success in destigmatizing victims of sexual violence is critical because it benefits not only female survivors, but also male survivors who might face even more barriers to speaking up due to social stereotypes (Davies and Rogers 2006). Consequently, we predict that this would further enhance consumers' aversion towards stereotypical, powerful portrayals of men and preference for more counter-stereotypical male narratives in films. Formally stated:

H_3: There is a positive effect of #MeToo on the performance of female-centric films that portray women using positive female counter-stereotypes.

H_4: There is a positive effect of #MeToo on the performance of male-centric films that portray men using positive male counter-stereotypes.

Collectively, we posit that #MeToo, through its collective action frames, creates a frame shift effect on consumers' film preference towards female (vs. male)-centric films, and such effect is further moderated by different gender portrayals in these films. Figure 1.1 summarizes our conceptual framework and core predictions.

FIGURE 1. 1. Conceptual Framework of Consumer Frame Shift in the Film Industry



While our theoretical framework predicts the action frame shift as the driver of #MeToo's effects, an alternative driver for #MeToo's effects can be argued as rooted in sexism. For example, consumers may interpret #MeToo as a threat to the status quo and therefore bolster gender stereotypes in defense of the existing system (Ellemers 2018; Yeung, Kay, and Peach 2014). Under such a prediction, we should observe stronger preferences for powerful male leads and weaker preferences for positive strong female leads (opposite of H2 and 3). Further, one could argue that the magnitude of sexual violence on women exposed by #MeToo could lead to a benevolent sexism effect, in which men are expected to provide for and protect women (Ellemers 2018; Jost and Kay 2005). Following that logic, we should observe a preference for the strong men protecting helpless women narrative, countering H1 and H2, or aversion towards the counter-stereotypical portrayals of men as gentle and caring (H4). In the following section, we test and provide empirical evidence to counter these alternative arguments.

Empirical Context and Strategy

Empirical Context and Data Construction

We began our investigation by collecting data on the top 150 performing films in each of the ten years from 2010 to 2019. During this time period, each year on average approximately 700 films were released in the US, the majority of which were small projects released for select audiences (Box Office Mojo 2021). Because #MeToo was a national movement, we selected a cut-off that included films with wide distributions. In addition, this time period (2010 to 2019) also signifies the massive growth of technology and the prevalence of social media underlying the proliferation of #MeToo (Lopes 2014).

We collected our data from three main sources: The Internet Movie Database (IMDb), Box Office Mojo (both owned by Amazon), and The-Numbers.com. These are credible, publicly available sources often referenced by industry stakeholders as well as academic researchers (Paulich and Kumar 2021; Ryoo, Wang, and Lu 2021). IMDb is one of the most comprehensive sources of film and TV series data, aggregating title, cast, crew, and studio details as well as crowdsourced user ratings and reviews. Box Office Mojo is a website that constantly tracks and updates box office data, providing information regarding daily sales and number of theaters of both domestic (US) and international releases. The-Numbers.com also collects these data and ranks film and star performance.

We first generated a list of the 150 top-performing films annually from 2010 to 2019 from The-Numbers.com, yielding a set of 1,500 films. We then matched these films with IMDb's academic database and used web-scraping programming to collect additional details regarding genres, principal cast and crew, release dates, plot summaries, keywords, etc. from the website. Using the IMDb's unique identifier, we collected films' financial information such as budget, number of theaters, and box office performance domestically and globally from Box Office Mojo. We then validated these financial details again with The-Numbers.com to maximize data accuracy. From this comprehensive list of films, we refined our dataset by excluding documentaries, animations, and non-English speaking movies, as these movies inherently activate different expectations, are processed differently, and are marketed and distributed for markets less directly affected by #MeToo's immediate cultural changes (Chow 2020; Pouliot and Cowen 2007).

For each film, we finally collected the details of its top four[1] credited principal cast and primary director/writer from IMDb's publicly available principal data set and coded the sex and race of each principal. We then used web-scraping programming to collect information regarding each actor's popularity in each year from 2010 to 2019 from The-Numbers.com. Our final dataset includes 3,494 unique principals (directors, writers, and actors) and their commercial popularity from 1,326 movies from 2010 to 2019. We do not include non-binary and transgender actors in our dataset due to their low frequency.

Empirical Approach

In this study, we investigated if and how #MeToo affected the performance of female (vs. male)-centric films. The eruption and virality of #MeToo on social media represented an exogenous shock to the motion picture industry, generating an appropriate context for a quasi-experimental design in which female (vs. male)-centric films were released either prior to or post #MeToo. We focused on studying the difference in box office performance between female- and

[1] IMDb's academic data set provides cast and crew details of its primary director, writer, and up to four top-credited cast per film.

male-centric films released before or after #MeToo. Subsequently, we investigated how this differential #MeToo effect was moderated by different gender portrayals of films' characters. We estimated these effects using a year and studio fixed-effects regression design.

One key threat to causal inference in this context is the endogenous effects of economic and societal trends that could both fuel the rise of #MeToo and affect film performance. For example, it could be possible that the growing prominence of gender equality issues and increasing divisiveness in the political environment causes #MeToo or similar movements, and influences public preference in films. Additionally, as streaming services with original content (e.g., Netflix, Amazon Video, etc.) become more popular over the years, these streaming services could be argued to have more prominent influences on theatrical film performance relative to social movements such as #MeToo. We used a year-fixed-effect design to address this issue and account for such industry-wide trends.

Another endogeneity concern lies in issues of film design, production, and overall quality. It is possible that following #MeToo, film producers invested more in female-centric films by employing more female directors, writers, and actors, indicating a simultaneous change in female-centric movie quality. However, our data set includes only the two years after #MeToo up to the SARS-CoV-2 pandemic. This timeframe ensures that any effects that we might observe are driven by a change in consumer preference, as opposed to a change in content production by studios since cast and plot decisions are made at least 2.5 years before theatrical release (Eliashberg, Elberse, and Leenders 2006; Eliashberg, Hui, and Zhang 2007; Follows 2018). Therefore, any film whose production was altered in response to #MeToo would have release dates that are outside of our time window (i.e., 2020 or later). Furthermore, this time frame ensures that the results are not contaminated by the effect of the global pandemic on the entertainment industry.

Additionally, it could be endogenously coincidental that films released post #MeToo had inherently better or worse quality. Studios (e.g., Disney, Sony, etc.) might have unique time-invariant characteristics that affect the resources, selection, production, and ultimately quality of released films. Even among films released by the same studio, the quality of the film might differ significantly by factors such as cast performance, storyline strength, etc. We attempted to minimize these issues by employing a studio and year fixed-effects model with a broad set of controls for film and cast quality and release strategy, in our models. To improve validity and balance between films released before and after #MeToo, we supplemented our analyses with a matching procedure. The following sections detail the variables used as well as the analyses and results.

Variables

In this section, we describe the variables used across our main models. For a complete set of all variables used in the main models and additional analyses, as well as their correlations, see Web Appendix A.

Dependent Variable

We are primarily interested in public reception of films centering on female (vs. male) characters. While there are multiple ways to measure film reception such as DVD sales (Ahmed and Sinha 2016; Mukherjee and Kadiyali 2011) or film ratings (Moon, Bergey, and Iacobucci 2010), box office performance is by far the most important metric for studios and consumers as it directly reflects market demand and consumer preference for films (Eliashberg, Elberse, and Leenders 2006). In this paper, we focused on *domestic US* box office performance (as opposed to *global* box office performance) due to #MeToo's origin and prevalence in the US. We obtained box office data from Box Office Mojo and validated with The-Numbers.com. We then adjusted all box office performances for inflation, log transformed the data to account for skewness, and use the transformed values as the dependent outcome across our models.

Independent Variables

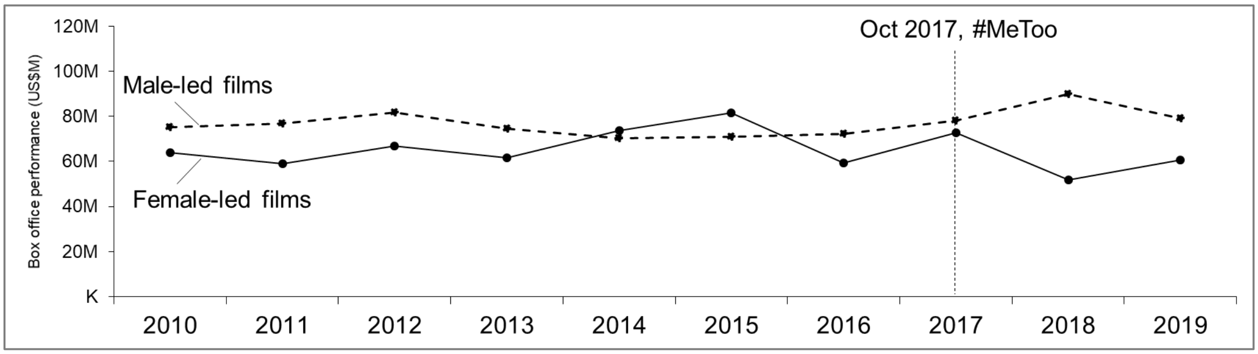
To address our research question, we focused on the interaction between whether a film is female (vs. male)-centric and whether it was released prior to or post #MeToo.

Female variable(s). There are multiple ways a film could be considered female-centric. First, it could be led by a female actor. This operationalization is most apparent to film audiences as the gender of the lead character is prominently shown on the film poster or the plot synopsis released by studios and used for advertising purposes. The lead cast is also the first billed or credited in the film credits, with the most screen time. We therefore coded dichotomous variable *f_lead* to indicate whether the first credited cast was female (vs. male)[2]. Second, a film centering on women might have more female representation in the cast. We operationalized this by calculating the proportion of female actors in the top four[3] actors of a film and treated this as a continuous variable *f_ratio* with value ranging from 0 (no one in the principal cast was female, e.g, *Dunkirk*) to 1 (all main cast were female, e.g., *Pitch Perfect*). Overall, 32% of films in our sample had a female lead, and the average female representation among main cast was 38%. Figure 1.2 demonstrates the model-free box office trend of female- and male-led films in our data set.

[2] Even for an ensemble cast, one cast is still most central to the story. For example, in *Avengers*, the first billed/credited cast is Robert Downey Jr. In the film poster, his visual is centrally featured in the largest size as compared to other stars).
[3]IMDb's academic data set provides cast and crew details of its primary director, writer, and up to four top-credited cast per film

Post (MeToo) variable. The #MeToo movement gained significant traction starting on Sunday, October 15, 2017, following a series of Tweets with the hashtag by prominent Hollywood celebrities as the Harvey Weinstein scandal broke out. We coded dichotomous variable *Post* to indicate if the film was released the weekend before (*Post* = 0) or after (*Post* = 1) the #MeToo movement became viral on social media.

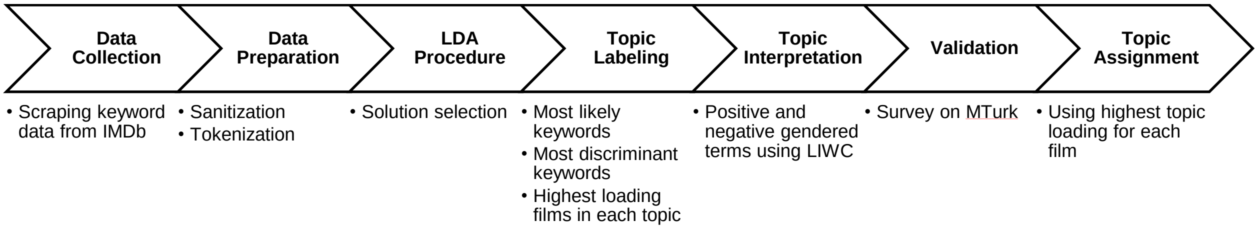
FIGURE 1. 2. Box Office Trends for Top Performing Films from 2010-2019



Gender Portrayals as Moderating Variables

In order to investigate the nuances in the portrayals of female and male characters in films, we applied a topic modeling approach to understand the film contents by analyzing film keywords collected from the IMDb. Figure 1.3 demonstrates our topic modeling procedure. This approach is beneficial for the current work for the following reasons. First, IMDb's keywords, covering any detail related to the content present in the film, were directly crowdsourced from viewers and validated by an IMDb editor. Past work often relies on studio-provided texts to analyze film contents (Berger, Kim, and Meyer 2021; Eliashberg, Hui, and Zhang 2007; Kuppuswamy and Younkin 2020; Paulich and Kumar 2021; for a review, see Web Appendix B). Compared to these approaches, for our current research question, using text analysis on crowdsourced keywords could uncover content themes as seen through the perspective of the audience instead of the film producers, providing more ecologically valid insights.

FIGURE 1. 3. Topic Modeling Process



Second, the use of topic modeling, while fuzzy in nature, is especially helpful in this context with a large textual corpus. Latent Dirichlet Allocation (LDA) topic modeling (Blei, Ng, and Jordan 2003) can discover patterns in the text that might be undetected using a pre-defined category classification approach (Humphreys and Wang 2018). When the data is rich and nuanced, deciding on a clear category (or topic) for a document based on pre-set definition becomes more challenging. The LDA technique alleviates this difficulty by grouping words that frequently occur together into topics while allowing for these words to be overlapped and texts to belong to multiple topics at varying degrees; researchers can then use this information to determine the most representative category for the text (Berger et al. 2020).

One key challenge of topic modeling lies in tokenization and data cleaning in which texts are divided into subsets of tokens (or in most cases, words) and prepared for analysis. To control for the same words in different forms (e.g., *hoped* and *hopes*) or synonymous words (e.g., *cellphone* and *mobile phones*), researchers often conduct stemming and lemmatization, reducing each word to its stem or lemma (Berger et al. 2020). This could, however, lead to over-stemming (i.e., different words reduced to the same stem: *university* and *universe* to *univers*) or under-stemming (i.e., words of the same root reduced to different stems: *alumni*, *alumnae*, *alumnus* to three different stems). In these aspects, the IMDb keywords have strong advantages over plot summaries or movie scripts, as they are already tokenized. Due to platform requirements, the keywords are all English words and phrases in non-capitalized letters, in singular form, and without accent (e.g., *café*). IMDb prompts keyword contributors to use existing keywords instead of creating new ones and the platform regularly monitors repetitive keywords (IMDb 2021). Therefore, using keywords instead of raw texts (e.g., scripts, summaries, or reviews) could significantly reduce the number of topics needed for LDA. Consequently, using LDA topic modeling on the IMDb keywords as tokens could result in both relevant and parsimonious topics.

Prior to performing topic modeling on the keywords, we excluded the keywords that had been controlled by other variables in our data set (e.g., adaptation or remake) and converted the keyword data into a document-term matrix, in which the set of all keywords in each film was a document, and each keyword was a term in that document. On average, a film in our data set had over 200 keywords; across 1,326 films, we obtained 54,725 unique keywords. The LDA procedure estimates terms that frequently occur together and groups them into topics with varying probabilities. The procedure then uses these term-level probabilities to calculate the distribution of a document over the generated topics. A film's set of keywords could be distributed over multiple topics with varying probabilities. Each keyword had a different likelihood of belonging to different topics. We performed multiple iterations of the LDA procedure to arrive at a solution that optimally balanced parsimony and discrimination among topics. The final solution consisted of three topics. Below we labeled and described these three topics, and in Web Appendices C and D we provided the complete detail of our process for developing and validating these topic labels and descriptions using predefined dictionaries (Linguistic Inquiry and Word Count — LIWC) and survey with over 600 participants.

TABLE 1. 2. Topic Description and Examples

Topic	Representative keywords	Topic description	Gender portrayals	Example film	Top keywords in film
Brutal	"violence," "pistol," "shotgun," "shot in the chest," "short to death," "held at gunpoint," "kidnapping," "blood," "blood splatter"	Highly violent theme. Powerful but isolated lead characters	*Female lead* Weak, victimized, vengeful, self-focused (negative stereotypes)	Peppermint (2019)	"revenge," "female vigilante," "murders," "murder spree," "brutal violence"
				Hanna (2011)	"female assassin," "child assassin," "revenge," "pistol," "shot to death
			Male lead Self-focused, violent, cold (negative stereotypes)	John Wick 3 (2019)	"assassin," "one-man army," "ultraviolence," "gun battle," "shot to death"
				Taken 3 (2014)	"tough guy," "murder," "death of wife," "one-man army," "kidnapping"
Aspirational	"explosion," "death," "supernatural," "good versus evil," "self-sacrifice," "friendship," "fantasy world," "fairy tale," "magic"	Empowering theme. Powerful, heroic and community-oriented lead characters	*Female lead* Powerful, heroic, other-focused (positive **counter-stereotypes**)	Star Wars IX (2019)	"female protagonist," "soldier," "rebel," "supervillain," "self-sacrifice"
				Wonder Woman (2017)	"matriarchy," "superheroine," "anti-war," "female scientist," "sister mother /daughter relationship"
			Male lead Powerful, heroic, other-focused (positive stereotypes)	Aladdin (2019)	"hero," "princess," "genie," "good versus evil," "prince," "good woman," "evil man"
				Harry Potter (2011)	"unlikely hero," "battle," "evil sorcerer," "friendship," "good wins," "strong female character"
Relational	"husband wife/ family relationships," "friendship," "telephone call," "kiss," "restaurant," "hospital"	Social and relationship-oriented themes Complex personal and psychological depiction of lead characters	*Female lead* gentle and warm, not powerful or violent (positive stereotypes)	Crazy Rich Asians (2018)	"rich man poor woman," "female professor," "elitism," "Asian American," "wedding ceremony"
				Me Before You (2016)	"parent son relationship," "birthday party," "working-class family," "male nurse," "unemployment," "accident"
			Male lead gentle and warm, not powerful or violent (positive **counter-stereotypes**)	Bohemian Rhapsody (2019)	"bisexual male," "music band," "singer," "rock concert," "playing piano," "friendship between men"
				The King's Speech (2010)	"period drama," "speech impediment," "king," "stuttering," "royal family," "world war two"

As summarized by Table 1.2, Topic *Brutal,* represented by keywords such as “violence,” “shot to death,” “shot in the chest,” “pistol,” etc., included films with highly violent films (e.g., *John Wick*, *Lucy*). Lead characters in this topic were powerful, vicious, and often alone (e.g., *John Wick*). Female characters in this topic fitted the damsel in distress archetype in films: they either needed protection (e.g. Lois Lane in *Men of Steel*) or were victimized and then turned vengeful (e.g. *Lucy*) while male characters were competent but cold and self-focused, signifying powerful male stereotypes. On the other hand, while also sharing some violence-related keywords, Topic *Aspirational* prominently features keywords that focus on positive and empowering themes such as “good versus evil,” “friendship,” “self-sacrifice,” “supernatural,” “fantasy world,” or “fairy tale.” Lead characters in this topic share positive traits of both genders: powerful and competent, yet warm and other-focused (e.g., Wonder Woman, Harry Potter). Thus, female characters in the *Aspirational* topic were counter-stereotypically powerful and competent. Finally, Topic *Relational* included keywords that emphasized relationship-oriented keywords such as “family/ husband wife/ father son/ mother son/ mother daughter relationship,” or “telephone call,” “kiss,” “restaurant,” “hospital,” etc. Lead characters in this topic were depicted as having complex psychological struggles and vulnerabilities. They exhibited the least stereotypical male traits, they were seen as the least powerful and violent, but rather warm and gentle, a depiction that is counter-stereotypical for male roles in films.

Finally, after validating these topics, we assigned each film to its highest loading topic and code *topic* as a categorical variable. Across our models, we used *topic* as the proxy variable for gender portrayals in films. We expected female-centric films and male-centric films in Topic *Brutal* to permit testing of H1 and 2, while female-centric films in Topic *Aspirational* to permit testing of H3 and male-centric films in Topic *Relational* to permit testing of H4.

Control Variables

In addition to the independent variables, we controlled for other explanatory variables that have been documented to impact box office performance. First, we included a series of film characteristic variables to control for differences in production detail and quality. Films with higher production budgets (e.g., blockbusters) tend to signal significant investment of time and quality by production studios. We collected budget information from Box Office Mojo, The-Numbers.com, and pre-release interviews by the production team and log transformed the budget to minimize skewness. For 13 independent films that did not have any publicly available budget information, we used the average production budget for a Sundance film as our best estimate (Bernstein 2015; excluding these films did not significantly alter the results). In addition, consumers might expect different types of films and quality from different studios (e.g., Disney). We therefore also used a categorical variable *studio* to code for seven major studios and their subsidiaries (see Web Appendix E for a complete list). Film format (IMAX or 3D) could signal better sound and visual quality and experience, and films released in IMAX or 3D might have higher ticket prices than other films. As a result, we coded variable *imax_3d* to denote if a film was released in either of these formats. To control for cast performance, cinematography, plot and script strength, and overall film quality, we included in our models the *meta_score* variable (ranging from 0 to 100) collected from Metacritic.com based on weighted evaluations by reputable experts and publications (one film did not have a *metacritic* score available at the time of data collection and was subsequently excluded from analyses).

In addition, certain film characteristics had been shown to have a significant impact on box office performance. Films that were adapted/remade from another creative work or based on historical/true events tend to perform better (Joshi and Mao 2012). Therefore, we added dichotomous variables *adapt*, deriving from the keywords collected for each film to account for this effect. In addition, films in a series (or with a sequel) often have a better chance at the box office (Basuroy and Chatterjee 2008; Heath et al. 2015). Consistent with prior research (Basuroy and Chatterjee 2008; Heath et al. 2015), we coded a categorical variable, *sequel*, to reflect whether a film was the first in a series (*sequel* =1; e.g., *The Hunger Games*), a sequel (*sequel* = 2; e.g., *The Hunger Games: Catching Fire*), or a stand-alone (*sequel* = 0; e.g., *Bohemian Rhapsody*). We included three dummy variables to reflect a film's MPAA rating (G, PG, PG-13, and R; no NC-17 films appeared in our dataset) as it could broaden or limit a film's appeal. We also created five dummy variables to control for the major genres (action, drama, comedy, thriller, and adventure) of films included in our sample.

Furthermore, films with a popular or powerful lead might attract more consumers (Elberse 2008). The variable *star_score* was collected from The-Numbers.com for each year from 2010 to 2019, which reflects the popularity and power of the lead actor in the respective year. *Star_score* is a composite measure of a star's commercial success in the respective year as well as the preceding two years. We also controlled for the race of the main lead with the dichotomous variable *lead_race*, which indicated whether the lead was non-white (*lead_race* = 1) or white (*lead_race* = 0). Dichotomous variable *Ensemble* indicated whether a film featured multiple principals who shared roughly similar screen time (e.g., *Love Actually*). In recent years, an ensemble cast usually consists of multiple powerful stars (e.g., *Avengers*), which could attract a larger audience.

Next, we controlled for film-release details. Categorical variable *release_wide* captured a film's release approach by the distribution studio. A studio might decide to first choose a limited release for a film in select venues to observe market reception and then send it to a wide release later (*release_wide* = 0). However, 82% of films in our sample were widely released (*release_wide* = 1) from the beginning, and a few films stayed in limited release during their entire run (*release_wide* = 2). Variable *re_release* indicated whether the film was being re-released by its studio. In some instances, older films were re-released when they had a newer version (e.g., a 3D format) or reached an important milestone (e.g., 20-year anniversary rerun of *Titanic*; excluding these re-released films did not change our results). Variable *max_theater* reflected the highest number of theaters the film was released to during its run time. The number of theaters impacts audience exposure to a film and also directly reflected the marketing budget spent on promoting the film (Paulich and Kumar 2021). In addition, seasonality (e.g., Christmas, summer, etc.) has a strong effect on box office performance (Einav 2007). To control for seasonality, we included dummy variables for the month the film was first released. We also added the *year* variable to indicate a film's US release year to account for time-varying factors that could affect box office performance across all films in theaters in the same year.

Analysis and Results

Differential Effects of #MeToo on Female- and Male-centric Films

Equation 1 estimates the effect of #MeToo on box office performance of female- and male-centric films using a studio and year fixed-effects model in which we primarily focused on the interaction term *Female×Post*. In this model, a film's box office performance is a function of the interaction term and its individual components, as well as the year and studio fixed-effects and other time-varying controls.

$$Y_{it} = \alpha_i + \beta_1 (\text{Female}_i \times \text{Post}_t) + \beta_2 \text{ Female}_i + \beta_3 \text{ Post}_t + \tau_t \text{Year} + \delta z_{it} + \varepsilon_{it} \qquad (1)$$

Aross our main models, we used Petersen's 2009 two-dimensional clustered standard errors to adjust for the possible correlations across studio and time. In this equation, ε_{it}

$= \gamma_i + \lambda_t + u_{it}$, where γ_i is the studio-specific component and λ_t is the time-specific component, and u_{it} pertains to idiosyncratic components unique to each observation. Additionally, since fixed-effects estimates, although more consistent, could be less efficient than random-effects estimators, we also ran random-effects models and used the Hausman test to validate the identification of the fixed-effects model. The Hausman test result ($\chi^2(8) = 55.61$, $p < .001$) indicated that the fixed-effects model best fit our context.

We estimated the above model across different operationalizations of female (vs. male)-centric films; results are reported in Table 1.3 and visualized in Web Appendix F). Consistently across both operationalizations of female-centric variables, the coefficient for *Female×Post* was negative and significant. We then explored the simple effects of #MeToo on female (vs. male)-centric films by reverse coding *Female* variables. Specifically, the coefficients for *Post* were significant and positive among male-led films or films with a higher proportion of male actors in the main cast ($\beta = .26$ and $.32$ respectively, *p*s <.05). However, among films that were female-led or had a higher proportion of female actors in the main cast, the coefficients for *Post* were not statistically significant. Together, these results suggested that post #MeToo, while male-centric films seemed to experience a boost, female-centric films did not experience the same effect at the ticket counters. However, the current result did not fully explain why and how this effect manifest. We therefore explored the nuances of the effect by investigating how audiences perceived different portrayals of female characters in these films.

TABLE 1. 3. Differential Effects of #MeToo on Box Office Performance of Female (vs. male)-centric Films

Studio fixed-effects model				
	DV: US Box Office Performance (log transformed)			
Variables	Model 1 Female=f_lead		Model 2 Female = f_ratio	
N = 1312	β	(SE)	β	(SE)
Female * Post	-.16 **	(.07)	-.30 **	(.14)
Female	.12 ***	(.04)	.14 *	(.07)
Post	.26 **	(.12)	.32 **	(.13)
Log (adj_bud)	.04 **	(.02)	.04	(.02)
meta_score	.01 ***	(.00)	.01 ***	(.00)
imax_3d	.04	(.06)	.04	(.06)
adapt	.00	(.03)	.00	(.03)
sequel=1	.59 ***	(.06)	.58 ***	(.06)
sequel=2	.26 ***	(.05)	.27 ***	(.05)
mpaa PG-13	-.19	(.19)	-.20	(.19)
mpaa PG	-.09	(.19)	-.11	(.19)
mpaa R-Rated	-.21	(.19)	-.22	(.19)
g_action	-.20 ***	(.04)	-.21 ***	(.04)
g_adventure	-.02	(.05)	-.02	(.05)
g_comedy	-.08 **	(.04)	-.08 **	(.04)
g_drama	-.03	(.04)	-.02	(.04)
g_thriller	-.02	(.04)	-.02	(.04)
star_score	.00 ***	(.00)	.00 ***	(.00)
Ensemble	.21 ***	(.06)	.21 ***	(.06)
lead_race	.16 ***	(.04)	.16 ***	(.04)
Release_wide=1	-.33 ***	(.06)	-.33 ***	(.06)
Release_wide=2	-.30 ***	(.08)	-.30 ***	(.08)
re_release	-.20	(.26)	-.20	(.26)
max_theater	.00 ***	(.00)	.00 ***	(.00)
R^2	.75		.75	

***p < .01; ** p < .05; * p < .1

Notes: Month and year dummies are omitted from the table due to limited space

Gender Portrayals in Films as Moderating Factor

Equation 2 estimates the #MeToo effect on box office performance of female (vs. male)-centric films as moderated by how gender were portrayed in these films using the same studio and year fixed-effects model. Gender portrayals in films were operationalized by categorical variable *Topic* with three levels; in this model we used Topic *Brutal* as the base level for comparison.

$$Y_{it} = \alpha_i + \beta_1 (Female_i \times Post_t \times Topic_i) + \beta_2 Female_i + \beta_3 Post_t + \beta_4 Topic_i + \beta_5 (Female_i \times Post_t) + \beta_6 (Female_i \times Topic_i) + \beta_7 (Topic_i \times Post_t) + \tau_t Year + \delta z_{it} + \varepsilon_{it} \quad (2)$$

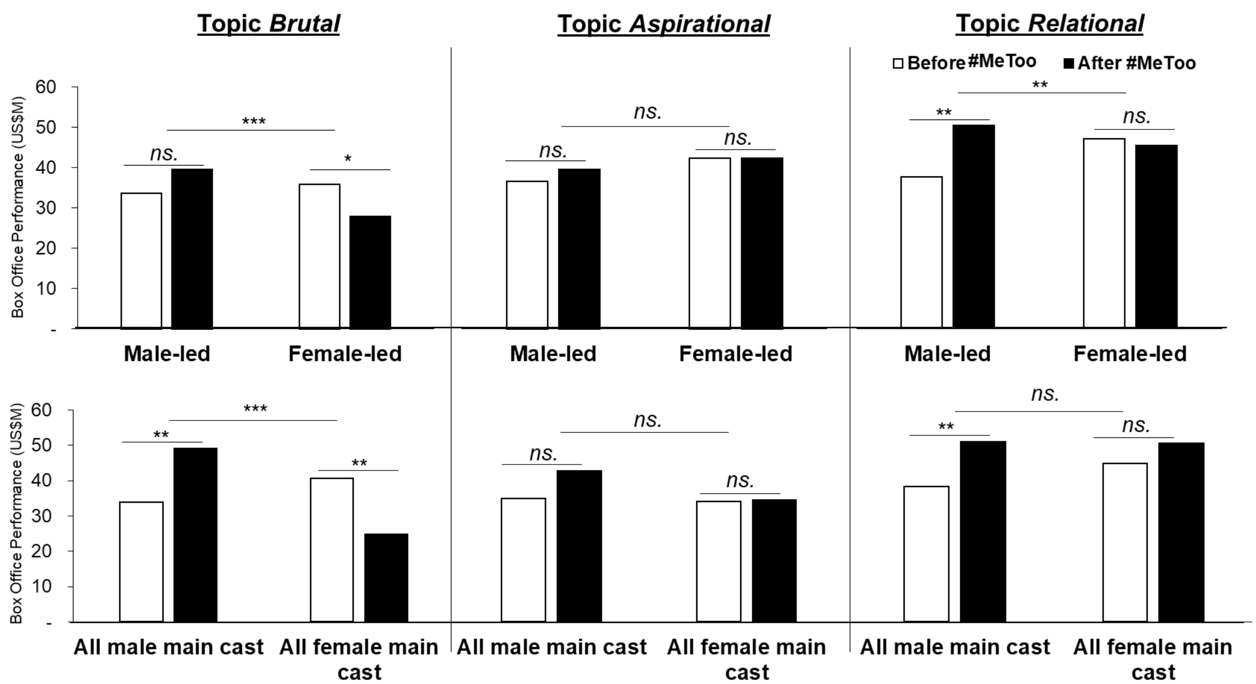
We are primarily interested in the three-way interaction of *Female×Post×Topic* to study if the treatment effect of #MeToo on films differed not only by the gender focus of the films but also by how the genders were depicted in these films. We again ran this model across different operationalizations of *Female*. Results are reported in Table 1.4 and visualized in Figure 1.4. Across models, we found significant three-way interactions between *Female, Post,* and *Topic* in consistent directions, indicating that effects of #MeToo on female- versus male-centric films differed significantly depending on how gender was portrayeds in these films. Figure 4 demonstrates the relationship between #MeToo and female- (vs. male)-centric films in each topic *Brutal, Aspirational,* and *Relational r*espectively.

TABLE 1. 4. Effects of #MeToo on Box Office Performance of Female (vs. Male)-centric Films, as Moderated by Female Portrayals.

Studio fixed-effects model				
	DV: US Box Office Performance (log transformed)			
Variables	Model 1 Female=f_lead		Model 2 Female = f_ratio	
N = 1312	β	(SE)	β	(SE)
Female × Post × topicAspirational	.46 **	(.22)	.80	(.43)
Female × Post × topicRelational	.21	(.19)	.84 **	(.38)
Female × Post	-.44 ***	(.17)	-1.00 ***	(.33)
Post × topicAspirational	-.08	(.12)	-.20	(.18)
Post × topicRelational	.13	(.10)	-.13	(.16)
Female × topicAspirational	.09	(.11)	-.22	(.23)
Female × topicRelational	.05	(.09)	-.03	(.19)
Post	.18	(.14)	.41 **	(.17)
Female	.07	(.08)	.19	(.16)
Topic=Aspirational	-.03	(.05)	.05	(.09)
Topic=Relational	.12 **	(.06)	.12	(.08)
R^2	.76		.75	

Notes: ***p<.001;**p<.01;*p<.05. Controls, Month and year dummies are omitted from the table due to limited space

FIGURE 1. 4. Differential Effects of #MeToo on Box Office Performance of Female (vs. Male)-centric Films



Notes: *** p < .01; ** p < .05; * p < .10; *ns.* not significant at .1 level. Box office values are back-transformed for interpretation, calculated at average values for all control variables.

We then examined the coefficients of *Female* × *Post* in each topic by recoding the base level of variable *Topic*. In topic *Brutal*, the coefficients for *Female* × *Post* were consistently negative and significant for female (vs. male)-led films (β = -.43, p = .01) and films with a higher proportion of females in the cast (β = -1.03, p = .002). In topic *Aspirational*, we observed no such significance or pattern across both operationalizations of *Female.* In Topic *Relational*, the coefficients for *Female* × *Post* were consistently negative but only significant for female (vs. male)-led films (β = -.23, p = .02). This suggests that the negative effect of #MeToo on female (vs. male)-centric films mainly manifests in Topics *Brutal* and *Relational*.

We further studied the simple slope of #MeToo on female (vs. male)-centric films in each of these topics by recoding the base level of *Topic* and *Female*. The results (as represented by Figure 1.4) showed that the negative differential #MeToo effect was mainly driven by a consistent negative impact on female-centric films in Topic *Brutal* and a (marginally) significant consistent positive impact on male-centric films in Topic *Relational*. We did not observe a significant effect of #MeToo on male-centric films in Topic *Aspirational* and *Brutal* or on female-centric films in Topics *Aspirational* and *Relational* (see Web Appendix F for detailed statistics on simple slopes).

Collectively, these results indicated a highly nuanced picture of the #MeToo effect on the film industry, reflecting the multifaceted nature of the movement as it unfolded in the marketplace. Consistent with H1, female-centric films in Topic *Brutal*, often depicting a negative albeit counter-stereotypically powerful female, were not well-received after #MeToo. Female-led films in this topic on average earned $8 million less after #MeToo . For example, *Peppermint* (2018), a film in Topic *Brutal*, featuring a victimized mother who turned murderous in her quest for justice, flopped at the box office. This, however, does not necessarily indicate a sexism effect in which female-centric films are punished and male-centric films are rewarded as a result of #MeToo. Although we did not find evidence for H2, male-centric films in Topic *Brutal* featuring stereotypical powerful male characters also did not receive a significant boost in box office performance after #MeToo. Rather, supporting H4, male-centric films that explore counter-stereotypical portrayals of men in topic *Relational* are rewarded. Male-led films in this topic earned $12 million more on average post #MeToo. For example, *Deadpool* 2 (2019), a male-led film in Topic *Brutal*, despite being part of a popular and successful series with a significant investment, did not perform as well as the previous film in the franchise. By contrast, *Bohemian Rhapsody* (2018), a film in Topic *Relational* that centered on the male lead’s vulnerability and personal struggles, received significant critical and financial success. Notably, in Topic *Aspirational* in which both male and female characters were powerful and positive, despite not getting a significant #MeToo boost, female-led films outperformed male-led ones ($\beta = .15$, $p = .04$). Taken together, the results support a frame-shift narrative in which after #MeToo, the public and consumers appreciated more well-rounded portrayals of men and women and disliked negative victimized portrayals of women in films.

Robustness Check

Matching as an alternative empirical approach. We first examined the robustness of our findings using propensity score matching as an alternative estimation approach. A matching procedure reduces bias in estimating treatment effects by identifying and comparing subjects in the treated group with their most similar match in the untreated group across a wide range of variables (Angrist, Imbens, and Rubin 1996). One of the most common matching methods is nearest neighbor propensity score matching (PSM), which matches treated subjects with their closest untreated counterpart based on propensity score (PS)—the likelihood that each subject would receive treatment based on a series of covariates (Morgan and Harding 2006). In our context, we strived to compare films released prior to #MeToo that were most similar to those released post #MeToo. Due to the nature of our sample, significantly more films were released prior to than post #MeToo, creating a favorable condition for PSM.

We first estimated the PS of films in our dataset to be released post #MeToo using a probit model with variable *Post* as the DV and a series of covariates as the IVs. These IVs were selected based on a series of backward stepwise probit regressions. Each film released after #MeToo was then matched with its closest counterpart among films released before #MeToo based on their PS. Subsequently, we check the quality of the matching procedure by calculating the average percentage reduction in bias (PRB) statistics across the variables. Overall, the average PRB across co-variates is 60%. To estimate the #MeToo effect, we ran a weighted OLS regression model on box office performance as a function *Female, Post*, and *Topic* and their interactions. The primary variable of interest was *Female*×*Post*×*Topic*. As shown in Table 1.5, we found significant three-way interactions in the same directions as in the main models for operationalization of female as *f_lead*. We observed directionally consistent and marginally significant results with variable female operationalized as *f_ratio*.

TABLE 1. 5. Matching Results

Weighted OLS with Matching Output				
	DV: US Box Office Performance (log transformed)			
Variables	Model 1 Female=f_lead		Model 2 Female = f_ratio	
N = 496	β	(SE)	β	(SE)
Female × Post × topicAspirational	1.72 ***	(.64)	.48	(.34)
Female × Post × topicRelational	1.30 **	(.56)	.51 *	(.30)
Female × Post	-1.24 **	(.50)	-.33	(.27)
Post × topicAspirational	-.97 ***	(.36)	-1.08 **	(.54)
Post × topicRelational	-.58	(.32)	-1.00 **	(.50)
Female × topicAspirational	-.42	(.48)	-.18	(.28)
Female × topicRelational	-.37	(.44)	-.01	(.25)
Post	.83 **	(.41)	.97	(.52)
Female	.31	(.38)	-.05*	(.22)
Topic=Aspirational	.53 *	(.27)	.63	(.42)
Topic=Relational	-.16	(.26)	-.23	(.40)
Intercept	17.36 ***	(.31)	17.52 ***	(.40)
R^2	.10		.10	

Notes: ***p<.01;**p<.01;*p<.05. Month and year dummies are omitted from the table due to limited space

Additional Analyses. For sensitivity analyses, we run the models using box office performance in the first month as our DV, and our results remain consistent and significant (Web Appendix G, Table 1W.9). Additionally, we also conduct exploratory analysis on the global box office of films in our data set. Although #MeToo is considered a global movement, we did not expect its effects to manifest globally in a similar manner, because different geographic regions might be subject to different cultural norms and #MeToo reached different countries at different times. Indeed, we did not find similar and significant results with the global box office performance of these films (Web Appendix G, Table 1W.9).

In categorizing films into topics from the topic modeling procedure, in the main models, we chose the topic that the film has the highest loadings on among the three topics. For sensitivity checks, we assigned a topic for a film only if the film had at least 50% loadings on the topic, ensuring that the topic dominantly represented the film (in doing this, we lost 14 films from our sample). Our results remained robust to this operationalization (Table 1W.10).

Furthermore, we used alternative ways to operationalize *Female* variable. First, films could have a female director and/ or writer that altered the perspective of gender portrayals in films. We coded dichotomous variable *F_rated* as 1 if the film was either directed or written by a woman, and 0 otherwise. Second, we also operationalized *Female* variable using the Bechdel test. To pass the Bechdel test, a film needs to have at least two women talking to each other about something other than a man ("Bechdel Test" 2019). We collected the Bechdel test results for films from Bechdeltest.com, a crowdsourced database for the Bechdel test, and coded the dichotomous *Bechdel* variable (1= pass, 0 = fail). In estimating equation (2), we found a significant and consistent effect of *F_rated×Post×Topic* on box office performance and a directionally consistent but not significant effect of *Bechdel×Post×Topic*, potentially because the Bechdel test result was not available for a significant portion of the films in our sample (see Web Appendix G, Table 1W.11).

Finally, we explored the intersectionality impact of #MeToo. Although not the focus of this research, coefficients on the *lead_race* variable are consistently significant across our models, consistent with recent findings by Kuppuswamy and Younkin (2020) that consumer preference is leaning towards a more diverse cast. We, however, did not find any significant interaction of our main variables with *lead_race*, suggesting that at least in the immediate two years following #MeToo, the effect was contained within its primary gender themes.

General Discussion

> "Social movements are critical to […] cultural shifts that make durable change possible […] Social movement actors changed the background against which questions of legality and justice were understood. They marched. They sang. They changed the wind."
>
> (Guinier and Torres 2013, p. 2743)

As modern social justice movements continue to evolve and permeate consumers' daily lives, it is critical to understand how such movements shape consumers' views and decisions. To our knowledge, the present research is the first to provide empirical evidence of how social justice movements can shape consumer spending decisions, even when the movements do not have such explicit aims. Through the study of the #MeToo movement and box office performance, we contend that social movements' collective action frames can reorient consumers' evaluations, thereby shaping their purchase behaviors. In this study, utilizing a uniquely obtained dataset and machine-learning techniques, we find that after #MeToo, female-centric films underperformed as compared to male-centric ones. Countering a sexism backlash explanation, we show that this effect occurred as a result of a decline in the performance of female-centric films that emphasized negative stereotypical portrayals of women (victimized, helpless) and a boost in the performance of male-centric films that featured counter-stereotypical portrayals of men (warm, empathetic). Critically, this pattern of effects emerged using varying operationalizations of female/male-centricity, different combinations of control variables, and alternative topic assignment approaches, as well as in a separate model that utilized propensity-score matching procedures. Further, because production studios generally finalize film production decisions 2.5 years before a film's release, our findings are not explained by endogenous production studio responses to #MeToo in the short run (i.e., within our sample window). Our findings generate important theoretical and managerial implications, as well as avenues for future research.

Theoretical Implication

Consumers both observe and participate in social justice movements (Nardini et al. 2021). While there has been ample research examining the effects of social movements on corporate and institutional practice (King and Pearce 2010) and legislative and policy work (Amenta et al. 2010), the effects of social justice movements on consumption are much less studied (Amenta and Polletta 2019). In this study, we address this gap and advance the marketing literature by bridging theories of collective action frames in sociology and framing effects in consumer psychology and demonstrating the spillover effects of social movements on consumer spending decisions. Specifically, we establish that social movements shift the frames against which consumers evaluate products (movies) and make purchase decisions. This further contributes to the burgeoning marketing strategy literature that aims to help market stakeholders who are navigating a quickly changing, politically charged, and activist-driven landscape (Bhagwat et al. 2020; Kotler and Sarkar 2018).

Through the study of #MeToo and the film industry, our work answers calls for more research in broader cultural and consumer-related impacts of social movements (Amenta and Polletta 2019; Nardini et al. 2021). Social movements primarily focus on enacting changes in policy and legislation (Amenta et al. 2010; Della Porta and Diani 2020). For such rule-shifting changes to be lasting, scholars argue, they have to be accompanied by culture-shifting changes (Guinier and Torres 2013; Stosodard 1997). In this work, we show how disruptive social movements such as #MeToo can undermine persistent (gender) stereotypes. Past work in stereotype adjustment shows how stereotypes could be updated or adjusted at the individual level (Ellemers 2018; Judd and Park 1993). Our work, therefore, extends the literature in stereotypes by demonstrating how stereotypes could be reoriented at the aggregate societal level.

Importantly, our findings add to the highly nuanced ways consumers react to social justice movements. We contrast our theoretical framework with the potential for a sexism reaction account in which the threat to the current power system posed by #MeToo prompted people to bolster existing gender stereotypes (Ellemers 2018). Following #MeToo, our work shows that instead of gravitating towards traditional gender roles and structures, consumers sought to understand and advocate for more sympathetic and less masculine portrayals of men and deviate from victimized negative portrayals of women. This demonstrates that consumers actively find ways to reconcile and address the underlying power imbalance of #MeToo, moving past the blaming and accusing narrative that opponents of the movement often criticize (McDonald 2019). Moreover, much literature in gender inequality focuses on raising awareness about prejudice against agentic women (i.e., women who acted more like men), as opposed to advocating for more acceptance towards men adopting stereotypically female traits (Eagly and Karau 2002; Rudman and Glick 2001). Such imbalance incidentally puts more expectations and responsibilities on modern women who struggle to balance both professional endeavors and caretaking (Bertrand, Kamenica, and Pan 2015). The #MeToo shifts demonstrate how the public gravitates to more balanced expectations of both genders: women should not have to depend on men or turn to extreme measures for protection, and men could be gentle, warm, and not physically powerful or violent. In doing so, we contribute to the literature by providing a parsimonious yet nuanced framework for understanding social justice movements and consumption that could be easily adopted and extended to other social movement contexts (e.g., Black Lives Matter, Gay Pride).

Managerial Implications

In the past, social movements' effects often lagged for years and decades, primarily because grassroots movements took longer to take shape and build a presence in the media (Amenta and Polletta 2019). The advancement of technology and social media has drastically accelerated this process, enabling social movements to dominate, influence, and reshape public opinion in a relatively short time (Della Porta and Diani 2020). For managers who are navigating this increasingly complex environment, we provide a framework for analyzing and understanding these social movements' effects on the consumption of many (cultural) products. Managers should first analyze the precipitating conditions in the industry, as well as the diagnostic, prognostic, and motivational framings of social movements to predict how these movements could reorient consumers in different ways, informing product design and promotion strategies that best reflect shifting cultural norms.

Our framework and findings have important implications for the $600 billion entertainment industry (Ballhaus and Chow 2021). For years, industry executives and content creators maintained that traditional gender stereotypes persisted in films and other content because they reflected market preference (Amenta and Polletta 2019; Ellemers 2018). Our work shows that disruptive social movements have the power to "change the wind" and shift consumer preference in a way that upends the industry's existing standards (Guinier and Torres 2013). For managers and producers who make casting and content decisions on a daily basis, our findings demonstrate that traditional (gender) stereotypes in content creation might not be the most fruitful or effective. Our findings reveal that female-led movies in topic *Brutal* on average earned approximately $8 million less after #MeToo, while male-led ones in topic *Relational* on average earned approximately $12 million more after #MeToo. Managers and content producers should carefully consider this cultural shift in delivering content that best caters to market preference. Indeed, after #MeToo, Lego committed to removing gender bias in toys (Russell 2021), while Disney moved forward with remakes of classics such as *Snow White* and *The Little Mermaid* in an attempt to change the gender narrative (Kroll 2021; N'Duka 2019). In addition, both male and female actors should also pay attention to these trends in evaluating the commercial potential of films and other content products.

Finally, as firms are reckoning with whether and how to engage in brand activism, a phenomenon that has become increasingly common in the market (Bhagwat et al. 2020), awareness of these cultural shifts could be of crucial value. Because brand activism is a risky strategy, a nuanced understanding of how social movements affect consumers could help reduce risk and provide useful guidance. Firms could apply our framework in analyzing social movements to their specific consumer base to optimize their branding and communication to be more relevant, thoughtful, and authentic.

Limitations and Future Research Direction

In this work, we attempt to understand social movements' effects on consumer preference through their movie-going decisions. While box office performance is an important indicator, our study does not cover the effect of #MeToo on streaming content on platforms such as Netflix, Hulu, HBO Max, etc. As streaming becomes more and more prevalent in our life, propelled further by pandemic-related interruptions to the movie theater experience, streaming performance could be crucial in measuring consumer shifts in light of social movements and a worthy investigation for future research.

Additionally, in this work, we describe the #MeToo effect as reflected at the aggregate level. Drawing on framing literature, we contend that #MeToo acts as a superframe that shifts consumers' evaluation of products and purchase decisions. Future research could delve into individual psychological processes that explain the #MeToo impact on consumers. For example, it could be that participation and support for the movement in person create a licensing effect in which consumers no longer feel the need to consciously support all the female-centric narratives in entertainment. We advocate for a more nuanced understanding of #MeToo at the individual level.

As with any major social movement, #MeToo also withstood significant internal and external challenges. #MeToo's collective action frames were met with counter-frames from opponents of the movement. For example, #HimToo was embraced by a minority of individuals to counter #MeToo, advocating for male victims of false allegations of sexual harassment, and reinforcing existing negative female stereotypes (Morris 2018). In our work, we did not find a negative effect of #MeToo on powerful male portrayals (H2) or a positive effect on positive powerful female portrayals (H3). It is possible that sexism had a countervailing effect on the impacts of #MeToo for these types of films, rendering the net effects of #MeToo null. Our findings in this study show that, at least immediately following #MeToo, the movement's primary collective action frames dominated the central stage of public opinion.

Finally, as a social movement moves past its peak in public discussion, one critical question is whether its changes are permanent. This remains an intriguing future research question. In the motion picture industry at least, we anticipate that the #MeToo effect could be enduring, primarily because #MeToo causes a structural change within the industry. Recent work by Luo and Zhang (2021) showed that #MeToo helped increase the hiring of female writers, directors, and other members of the production teams. We expect that such fundamental changes will reinforce the #MeToo effect in the industry by delivering higher quality films that portray more nuanced and balanced gender dynamics, enhancing the performance of female-centric films in the long run. Because films as cultural products not only reflect dominant social values but also set social standards (Amenta and Polletta 2019; Eliashberg, Elberse, and Leenders 2006), we believe that the ripple effect of #MeToo could reach consumers in generations to come.

> "Major social movements eventually fade into the landscape not because they have diminished but because they have become a permanent part of our perceptions and experience."
>
> (Adler 1975, p.5)

Web Appendix

Web Appendix A: Variable Summary and Correlations

TABLE 1W. 1. Summary of All Variables

	Variable	Description	Source	N	Mean	SD	Med	Skew
	Dependent Variables							
1	adj_us	US box office performance in 2019 value, in US$M	BOM[1], TN[2]	1326	71.97	95.54	41.85	3.64
2	adj_w4	US box office performance in 2019 value, first four weeks in US$M	BOM	1326	63.17	85.00	36.98	3.80
3	adj_ww	worldwide box office performance in 2019 value, in US$M	TN	1326	18.20	276.43	81.63	3.62
	Independent Variables and Moderators							
	Female							
4	f_lead	if the film's lead is female (=1), or male (=0)	IMDb,M[3]	1326	.32	.47	0	.78
5	f_ratio	proportion of female actors in the main cast	IMDb, M	1326	.38	.23	.25	.49
6	f_rated	if the film's writer or director is female (=1), or male (=0)	IMDb, M	1326	.14	.35	0	2.09
7	bechdel	if the film passes the bechdel test (=1); fails (=0)	BD[4]	1166	.58	.49	1	-.33
8	*Post*[5]	released before (=0) or after (=1) #MeToo	IMDb	1313	.22	.41	0	1.3
	Topic		IMDb Keywords					
9	Brutal	films with highest loadings on topic *Brutal*		1326	.27	.45	0	1.01
10	Aspirational	films with highest loadings on topic *Aspirational*		1326	.23	.42	0	1.26
11	Relational	films with highest loadings on topic *Relational*		1326	.49	.50	0	.03
	Top		IMDb Keywords					
12	Brutal	films with at least 50% loadings on topic *Brutal*		1326	.27	.44	0	1.03
13	Aspirational	films with at least 50% loadings on topic *Aspirational*		1326	.23	.42	0	1.27
14	Relational	films with at least 50% loadings on topic *Relational*		1326	.49	.50	0	.05
	Controls							
	Film characteristics							
15	adj_bud	adjusted production budget	BOM, Wiki	1326	53.20	58.98	31.58	1.91
16	studio[6]	parent studio in the US	BOM	1326				
17	imax_3d	if the film has either imax or 3D format (=1; 0 otherwise)	IMDb	1326	.07	.25	0	3.43
18	meta_score	film score as rated by metacritic.com (0 – 100)	Metacritic	1325	54.35	17.57	53	.48
19	adapt	if the film was adapted or remade from other work (=1; 0 otherwise)	IMDb	1326	.48	.50	0	.10
20	sequel	if the film is not part of a series (=0)	BOM	1326	.81	.70	0	1.83
21		if the film is first in a series (=1)			.06			
22		If the film is a sequel in a series (=2)			.13			
23	mpaa	MPAA rating of the film (G or unrated = 0)	IMDb	1326	.01			
24		if the film is rated PG (=1)			.10			
25		if the film is rated PG-13 (=2)			.47			
26		if the film is rated R (=3)			.42			

Notes: [1]Box Office Mojo (BOM); [2]The-Numbers.com (TN); [3]Manually coded by RA (M), [4]Bechdel.com (BD); [5] #MeToo went viral during the weekend of October 15, 2017 (Sunday), 14 films released on that weekend were not coded for variable *Post*. Including these films as being released before #MeToo did not change the significance and direction of any effect found across models. [6]Details for the studio can be found in Web Appendix D

… continued Table 1W.1.

	Variable	Description	Source	N	Mean	SD	Med	Skew
	Genre							
27	g_action	if the film genre includes action (=1; 0 otherwise)	IMDb	1326	.33	.47	0	.71
28	g_adventure	if the film genre includes adventure (=1; 0 otherwise)	IMDb	1326	.23	.42	0	1.30
29	g_comedy	if the film genre includes comedy (=1; 0 otherwise)	IMDb	1326	.35	.48	0	.65
30	g_drama	if the film genre includes drama (=1;1 otherwise)	IMDb	1326	.53	.50	1	-.11
31	g_thriller	if the film genre includes thriller (=1; 0 otherwise)	IMDb	1326	.18	.38	0	1.68
	Cast power							
32	star_score	star score in respective year, as calculated by TN (0 – 561)	TN	1326	102.58	105.98	75.50	1.24
33	ensemble	if the film has an ensemble cast (=1; 0 otherwise)	IMDb	1326	.07	.26	0	3.36
34	lead_race	if the lead is a minority (=1; 0 otherwise)	M	1326	.18	.39	0	1.66
	Release details							
35	release_wide	if the film was first in limited release, then wide release (=0)	BOM	1326	.12			
36		if the film was first in wide release (=1)			.82			
37		if the film was only in limited release (=2)			.06			
38	re_release	if the film was a re_release (=1) or an original one (=0)	BOM	1326	.00	.05	0	18.10
39	max_theater	the maximum number of theaters in the film's run	BOM	1326	2,650	1,027	2,882	-.71
40	month	the month of film release (seasonality) Jan = 1	BOM	1326	.07			
41		February = 2			.07			
42		March = 3			.09			
43		April = 4			.07			
44		May = 5			.07			
45		Jun = 6			.09			
46		July = 7			.06			
47		August = 8			.10			
48		September = 9			.08			
49		October = 10			.09			
50		November = 11			.10			
51		December = 12			.10			
52	year	the year of the film's US release (base = 2010)	BOM	1326	.10			
53		2011			.10			
54		2012			.10			
55		2013			.10			
56		2014			.10			
57		2015			.10			
58		2016			.10			
59		2017			.10			
60		2018			.10			
61		2019			.10			

Notes: [1]Box Office Mojo (BOM); [2]The-Numbers.com (TN); [3]Manually coded by RA (M), [4]Bechdel.com (BD); [5]Details for the studio can be found in Web Appendix D

TABLE 1W. 2. Correlations of All Variables Used

	Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21
1	adj_us	**1.00**																			
2	adj_w4	**.99**	**1.00**																		
3	adj_ww	**.91**	**.91**	**1.00**																	
4	f_lead	**-.06**	**-.06**	**-.10**	**1.00**																
5	f_ratio	**-.12**	**-.12**	**-.15**	**.48**	**1.00**															
6	f_rated	-.04	-.05	**-.06**	**.40**	**.29**	**1.00**														
7	bechdel	**.07**	**.06**	.05	**.46**	**.42**	**.26**	**1.00**													
8	Post	.01	.01	.01	**.10**	.05	**.17**	.03	**1.00**												
9	topic_Brutal	**.16**	**.17**	**.17**	**-.19**	**-.22**	**-.13**	**-.12**	-.05	**1.00**											
10	topic_Aspirational	**.14**	**.15**	**.18**	-.03	**-.08**	-.03	-.04	-.01	**-.34**	**1.00**										
11	topic_Relational	**-.26**	**-.27**	**-.30**	**.19**	**.26**	**.14**	**.14**	.05	**-.60**	**-.54**	**1.00**									
12	top_Brutal	**.16**	**.17**	**.18**	**-.19**	**-.22**	**-.13**	**-.12**	-.04	**1.00**	**-.34**	**-.60**	**1.00**								
13	top_Aspirational	**-.26**	**-.28**	**-.31**	**.19**	**.27**	**.14**	**.14**	.05	**-.60**	**-.54**	**1.00**	**-.60**	**1.00**							
14	top_Relational	**.14**	**.15**	**.18**	-.03	**-.08**	-.03	-.03	-.01	**-.34**	**1.00**	**-.54**	**-.34**	**-.54**	**1.00**						
15	adj_bud	**.69**	**.71**	**.75**	**-.19**	**-.21**	**-.09**	**-.08**	-.05	**.24**	**.29**	**-.47**	**.25**	**-.47**	**.30**	**1.00**					
17	imax_3d	**.20**	**.21**	**.25**	**-.06**	**-.07**	**-.06**	-.04	**-.14**	**.06**	**.17**	**-.20**	**.07**	**-.20**	**.17**	**.36**	**1.00**				
18	meta	**.18**	**.14**	**.15**	.00	-.03	**.10**	**.06**	**.06**	**-.10**	-.05	**.13**	**-.10**	**.13**	-.05	-.01	**-.09**	**1.00**			
19	adapt	**.15**	**.13**	**.16**	-.02	**-.08**	.03	-.05	.02	-.02	**.15**	**-.11**	-.02	**-.11**	**.15**	**.21**	**.12**	**.13**	**1.00**		
20	sequel_0	**-.44**	**-.45**	**-.45**	**.08**	**.10**	.04	.00	-.01	**-.18**	**-.06**	**.20**	**-.18**	**.21**	**-.06**	**-.38**	**-.17**	.05	.01	**1.00**	
21	sequel_1	**.26**	**.26**	**.22**	**-.06**	**-.06**	.01	-.02	-.01	.04	.01	-.05	.04	-.05	.01	**.14**	.04	.03	.04	**-.50**	**1.00**
22	sequel_2	**.33**	**.34**	**.38**	-.05	**-.07**	-.05	.02	.02	**.17**	**.06**	**-.20**	**.18**	**-.21**	**.06**	**.34**	**.17**	**-.08**	-.04	**-.81**	**-.09**
23	mpaa_no_0	-.05	-.05	-.03	.00	.01	.02	.00	.05	.01	.04	-.04	.01	-.04	.04	-.05	-.02	.00	-.03	-.01	-.02
24	mpaa_no_1	.00	.00	.00	.04	.05	.02	.05	.03	**-.16**	**.20**	-.02	**-.16**	-.03	**.20**	.03	**.06**	**-.08**	**.09**	-.01	.01
25	mpaa_no_2	**.21**	**.22**	**.24**	.01	.00	.02	.03	**-.06**	.00	**.11**	**-.09**	.00	**-.09**	**.10**	**.29**	**.09**	**-.13**	**.09**	**-.09**	.04
26	mpaa_no_3	**-.21**	**-.21**	**-.23**	-.04	-.04	-.04	**-.06**	.03	**.10**	**-.24**	**.11**	**.10**	**.11**	**-.24**	**-.31**	**-.12**	**.18**	**-.14**	**.11**	-.04
27	g_action	**.27**	**.28**	**.31**	**-.22**	**-.26**	**-.12**	**-.18**	-.03	**.62**	.05	**-.59**	**.62**	**-.60**	.05	**.50**	**.17**	**-.16**	**.06**	**-.25**	**.08**
28	g_adventure	**.42**	**.43**	**.49**	**-.11**	**-.16**	-.04	-.03	.03	**.21**	**.27**	**-.42**	**.21**	**-.43**	**.28**	**.63**	**.25**	-.04	**.16**	**-.32**	**.12**
29	g_comedy	**-.11**	**-.11**	**-.16**	.02	**.13**	.03	.04	-.01	**-.17**	**-.15**	**.27**	**-.17**	**.28**	**-.15**	**-.18**	**-.11**	**-.07**	**-.22**	.04	-.01
30	g_drama	**-.25**	**-.27**	**-.26**	**.11**	**.11**	**.11**	.05	.03	**-.29**	**-.07**	**.31**	**-.29**	**.32**	**-.07**	**-.34**	**-.17**	**.32**	**.13**	**.33**	**-.18**
31	g_thriller	**-.07**	**-.06**	**-.06**	.01	-.01	**-.07**	-.03	-.03	**.21**	**-.10**	**-.11**	**.21**	**-.10**	**-.10**	**-.11**	-.02	**-.08**	**-.10**	-.05	.04
32	star_score	**.43**	**.43**	**.39**	**-.16**	**-.13**	**-.10**	-.05	-.03	**.21**	.00	**-.19**	**.22**	**-.19**	.00	**.40**	.02	.02	.03	**-.20**	.04
33	ensemble	**.26**	**.27**	**.27**	-.03	**-.06**	-.01	.06	.00	.05	.03	**-.07**	.05	**-.07**	.03	**.22**	**.06**	**.06**	.03	**-.13**	**.11**
34	lead_race	-.03	-.03	-.05	-.02	.05	.05	-.01	**.09**	-.02	-.04	.05	-.01	.05	-.05	**-.08**	.00	-.02	**-.08**	-.01	.02
35	release_wide_0	**-.13**	**-.18**	**-.13**	.04	**.05**	**.09**	.04	.01	**-.16**	**-.08**	**.21**	**-.16**	**.21**	**-.08**	**-.19**	**-.07**	**.42**	**.08**	**.16**	**-.06**
36	release_wide_1	**.21**	**.26**	**.19**	-.05	**-.07**	**-.11**	-.04	.02	**.21**	**.12**	**-.29**	**.21**	**-.29**	**.12**	**.27**	**.10**	**-.51**	**-.06**	**-.21**	**.09**
37	release_wide_2	**-.17**	**-.17**	**-.13**	.02	.04	**.06**	.02	-.05	**-.12**	**-.09**	**.18**	**-.13**	**.18**	**-.09**	**-.17**	**-.07**	**.25**	-.02	**.11**	**-.06**
38	re_release	-.02	-.01	**.20**	-.01	-.02	-.02	.05	-.03	-.03	**.07**	-.03	-.03	-.05	**.09**	.00	-.01	.05	.03	-.04	-.01
39	max_theater	**.57**	**.59**	**.54**	**-.09**	**-.11**	**-.09**	-.01	.05	**.27**	**.20**	**-.41**	**.27**	**-.41**	**.20**	**.61**	**.19**	**-.25**	.05	**-.38**	**.17**
40	mon_no_1	**-.08**	**-.08**	**-.08**	.00	-.04	-.05	-.03	-.04	**.06**	-.04	-.02	**.06**	-.02	-.04	**-.07**	-.02	**-.16**	**-.05**	.01	-.01
41	mon_no_2	-.03	-.02	-.03	.00	-.03	-.03	-.03	-.02	.00	.04	-.03	.00	-.03	.04	-.02	-.01	**-.10**	.01	.05	-.02
42	mon_no_3	-.02	-.01	-.01	.05	.04	.04	.04	-.01	.02	.02	-.03	.02	-.03	.02	.00	.00	**-.07**	.01	.01	.05
43	mon_no_4	.00	.00	.01	.00	.02	-.05	.03	-.02	-.01	-.04	.05	-.01	.04	-.04	-.04	-.03	**-.07**	-.01	.00	-.03
44	mon_no_5	**.11**	**.12**	**.12**	.00	.02	.05	.03	.00	-.02	.05	-.02	-.02	-.02	.05	**.16**	.04	.00	-.02	**-.06**	.01
45	mon_no_6	**.06**	**.07**	**.05**	.00	.01	.00	.01	-.01	.03	-.01	-.02	.03	-.02	-.01	**.06**	.03	.03	-.03	**-.06**	.02
46	mon_no_7	**.05**	**.06**	**.06**	-.03	.02	-.02	-.01	-.05	.03	.00	-.03	.03	-.03	.00	**.08**	**.06**	**.08**	-.02	-.05	.02
47	mon_no_8	**-.06**	**-.06**	**-.08**	.01	-.01	.00	-.01	.01	.05	-.01	-.04	.05	-.03	-.01	**-.07**	-.01	-.04	.00	.02	-.01
48	mon_no_9	**-.07**	**-.07**	**-.08**	.01	.01	-.03	-.02	-.03	.04	-.05	.01	.04	.01	-.05	**-.10**	.00	.01	.00	.02	.02
49	mon_no_10	**-.07**	**-.06**	**-.05**	-.03	-.02	-.02	-.02	.02	-.03	-.01	.04	-.04	.04	-.01	**-.08**	.01	.04	.01	-.01	.02
50	mon_no_11	.02	.01	.02	-.02	-.01	.03	.02	**.08**	**-.07**	.00	**.06**	**-.06**	**.06**	-.01	.01	-.05	**.14**	.05	.03	-.03
51	mon_no_12	**.09**	**.06**	**.06**	.00	-.01	.05	.00	.04	**-.08**	.05	.03	**-.07**	.02	.05	**.08**	-.01	**.11**	.04	.01	-.03
52	year_2010	.00	.00	-.02	.01	.00	-.03	.03	**-.18**	.01	-.02	.01	.00	.01	-.01	.04	.03	-.05	.00	.03	.02
53	year_2011	.00	.00	-.01	-.03	.01	-.05	-.03	**-.17**	-.03	.01	.02	-.03	.02	.00	.02	**.08**	-.02	.02	**.05**	-.02
54	year_2012	.02	.02	.04	-.03	-.05	-.02	-.03	**-.18**	.00	-.01	.01	.00	.01	-.01	.02	.03	.00	-.01	-.02	.05
55	year_2013	.00	.00	.00	-.05	-.02	**-.06**	.01	**-.18**	**.07**	-.03	-.04	**.07**	-.05	-.02	.02	.05	-.01	.00	-.01	-.02
56	year_2014	-.01	-.01	-.01	**-.07**	-.03	**-.05**	-.05	**-.18**	.00	.04	-.03	-.01	-.03	.04	-.02	.04	-.01	-.01	-.04	.03
57	year_2015	.00	.00	.00	-.01	.00	-.04	-.01	**-.18**	.01	-.05	.03	.01	.03	-.05	-.02	.00	.00	-.03	.02	-.01
58	year_2016	-.01	-.01	-.01	.01	.02	-.01	.04	**-.18**	.00	.05	-.04	.00	-.04	.05	.01	-.01	.00	-.01	-.02	-.03
59	year_2017	.01	.01	.02	.02	.01	.02	.04	.00	.00	.01	-.01	.00	-.01	.02	.02	-.05	.02	.01	.00	-.01
60	year_2018	.00	.00	.01	**.09**	.04	**.12**	.01	**.64**	.01	-.04	.03	.01	.03	-.05	-.03	**-.08**	.04	.04	.02	-.02
61	year_2019	-.01	-.01	-.01	**.06**	.03	**.12**	.00	**.62**	**-.06**	.04	.02	**-.06**	.02	.04	-.05	**-.09**	.03	.00	-.03	.00

Notes: Boldfaced values for correlations indicate significance at a 95% level.

…continued Table W2

Variables	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
1 adj_us																				
2 adj_w4																				
3 adj_ww																				
4 f_lead																				
5 f_ratio																				
6 f_rated																				
7 Bechdel																				
8 Post																				
9 topic_Brutal																				
10 topic_Aspirational																				
11 topic_Relational																				
12 top_Brutal																				
13 top_Aspirational																				
14 top_Relational																				
15 adj_bud																				
17 imax_3d																				
18 meta																				
19 adapt																				
20 sequel_0																				
21 sequel_1																				
22 sequel_2	**1.00**																			
23 mpaa_no_0	.02	**1.00**																		
24 mpaa_no_1	.01	-.03	**1.00**																	
25 mpaa_no_2	**.08**	**-.08**	**-.32**	**1.00**																
26 mpaa_no_3	**-.09**	**-.07**	**-.29**	**-.80**	**1.00**															
27 g_action	**.24**	.04	**-.14**	**.17**	**-.08**	**1.00**														
28 g_adventure	**.29**	.02	**.11**	**.22**	**-.29**	**.48**	**1.00**													
29 g_comedy	-.04	.00	**.10**	**-.11**	.04	**-.23**	**-.15**	**1.00**												
30 g_drama	**-.27**	-.01	-.02	-.01	.02	**-.38**	**-.36**	**-.16**	**1.00**											
31 g_thriller	.03	.03	**-.16**	**-.07**	**.16**	**.06**	**-.17**	**-.33**	**-.18**	**1.00**										
32 star_score	**.21**	**-.06**	**-.08**	**.15**	**-.09**	**.25**	**.18**	.05	**-.16**	-.05	**1.00**									
33 ensemble	**.07**	.01	-.04	**.07**	-.05	.03	**.08**	.04	-.05	**-.06**	**.07**	**1.00**								
34 lead_race	.01	**.08**	.00	.02	-.04	.02	-.02	-.02	.03	-.02	.00	-.04	**1.00**							
35 release_wide_0	**-.15**	-.03	**-.07**	**-.09**	**.14**	**-.20**	**-.15**	.00	**.31**	**-.07**	**-.09**	-.03	-.02	**1.00**						
36 release_wide_1	**.18**	**-.10**	**.08**	**.13**	**-.16**	**.25**	**.18**	-.02	**-.37**	**.10**	**.19**	.04	.00	**-.79**	**1.00**					
37 release_wide_2	**-.09**	**.21**	-.03	**-.09**	**.07**	**-.13**	**-.08**	.03	**.18**	**-.07**	**-.18**	-.02	.03	**-.09**	**-.53**	**1.00**				
38 re_release	**.06**	**.16**	.03	.00	-.05	.02	**.07**	-.04	-.03	-.03	-.02	**.09**	-.03	-.02	-.01	.04	**1.00**			
39 max_theater	**.32**	**-.11**	**.06**	**.20**	**-.22**	**.38**	**.41**	-.05	**-.45**	.04	**.43**	**.13**	**-.09**	**-.40**	**.68**	**-.55**	-.03	**1.00**		
40 mon_no_1	-.01	-.02	.02	-.03	.02	**.08**	-.01	-.03	-.05	**.07**	-.01	**-.06**	.01	**-.07**	**.09**	-.04	-.01	.01	**1.00**	
41 mon_no_2	-.04	.01	-.02	.04	-.03	.04	-.02	-.01	-.02	**.06**	-.01	**-.06**	.02	**-.09**	**.10**	-.03	.04	.04	**-.08**	**1.00**
42 mon_no_3	-.04	-.03	.02	-.02	.01	.00	.01	-.02	.02	-.01	-.04	.00	-.03	-.02	.00	.02	.03	-.04	**-.09**	**-.09**
43 mon_no_4	.02	-.02	.01	.04	-.04	-.03	-.03	.00	-.04	-.01	-.05	.00	.03	**-.07**	**.07**	-.01	.04	.01	**-.08**	**-.08**
44 mon_no_5	**.06**	.01	.04	.05	**-.07**	.02	**.09**	**.09**	**-.09**	**-.06**	**.06**	.02	-.01	-.04	.03	.02	-.02	**.08**	**-.08**	**-.08**
45 mon_no_6	.05	-.03	-.03	.00	.03	.05	**.06**	.05	**-.08**	-.05	**.07**	.01	-.02	-.04	.00	**.06**	-.02	.03	**-.08**	**-.09**
46 mon_no_7	.04	.05	-.02	.02	-.02	.05	**.05**	.03	-.04	-.01	.02	.02	-.03	.01	-.02	.01	-.01	.04	**-.07**	**-.07**
47 mon_no_8	-.01	-.03	.04	-.01	-.01	.04	-.01	.03	-.02	**.06**	**-.08**	.02	.03	**-.05**	**.08**	-.05	-.02	.00	**-.09**	**-.09**
48 mon_no_9	-.04	.05	-.03	.00	.02	-.04	-.05	**-.05**	.04	.04	-.02	.01	.00	.01	.00	-.01	.04	.00	**-.08**	**-.08**
49 mon_no_10	.00	-.03	-.01	**-.07**	**.08**	**-.06**	**-.08**	-.05	.05	**.07**	**-.07**	**-.06**	-.02	.00	-.01	.02	-.02	-.05	**-.09**	**-.09**
50 mon_no_11	-.01	-.03	-.02	.01	.01	**-.06**	-.01	-.04	**.13**	**-.06**	**.06**	.02	.03	**.12**	**-.10**	-.02	-.02	-.05	**-.09**	**-.09**
51 mon_no_12	.00	**.06**	.01	-.01	-.01	**-.07**	-.01	.01	**.08**	**-.07**	**.07**	**.06**	.00	**.21**	**-.20**	.03	-.02	-.05	**-.09**	**-.09**
52 year_2010	-.04	.00	**.08**	-.03	-.01	-.02	-.04	.03	-.03	.00	.00	.03	-.01	-.05	.02	.04	-.02	-.01	.02	-.01
53 year_2011	-.05	-.03	.01	.04	-.05	-.01	-.02	.04	-.05	.02	.04	-.01	**-.06**	**-.09**	**.05**	.04	-.02	.00	-.02	.00
54 year_2012	-.01	-.03	-.01	.01	.00	.01	.02	.02	-.03	.00	.02	.02	**-.06**	-.03	.01	.03	**.07**	-.02	.00	.03
55 year_2013	.02	.03	**-.08**	.03	.02	.01	-.03	-.01	-.02	**.07**	.04	-.01	.00	**.07**	-.04	-.03	**.07**	-.03	.01	-.01
56 year_2014	.03	-.03	.02	.00	.00	.02	-.02	-.01	.02	.02	-.04	-.01	-.01	.01	-.03	.03	-.02	-.03	-.01	.01
57 year_2015	-.02	.00	-.01	.00	.01	-.02	.01	-.01	.01	-.02	.00	-.03	-.02	**.06**	-.03	-.03	-.02	.00	.02	-.01
58 year_2016	.04	.00	-.03	.03	-.01	.03	.02	.02	.02	-.05	-.04	.04	.00	-.02	.02	.00	-.02	.01	-.01	.02
59 year_2017	.00	.00	-.02	.00	.01	.01	.01	-.05	.05	-.01	.03	-.01	.03	.05	-.03	-.02	-.02	.01	.01	-.02
60 year_2018	-.01	-.03	.01	-.02	.02	.02	.04	-.02	-.02	.00	-.05	-.03	**.06**	.00	.03	-.04	-.02	.04	-.01	-.02
61 year_2019	.04	**.07**	.03	-.04	.01	-.04	.00	-.01	.04	-.03	-.02	.02	**.07**	.00	.02	-.03	-.02	.03	-.02	.01

Notes: Boldfaced values for correlations indicate significance at a 95% level.

…continued table W2

	Variables	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
1	adj_us																				
2	adj_w4																				
3	adj_ww																				
4	f_lead																				
5	f_ratio																				
6	f_rated																				
7	bechdel																				
8	Post																				
9	topic_Brutal																				
10	topic_Aspirational																				
11	topic_Relational																				
12	top_Brutal																				
13	top_Aspirational																				
14	top_Relational																				
15	adj_bud																				
17	imax_3d																				
18	meta																				
19	adapt																				
20	sequel_0																				
21	sequel_1																				
22	sequel_2																				
23	mpaa_no_0																				
24	mpaa_no_1																				
25	mpaa_no_2																				
26	mpaa_no_3																				
27	g_action																				
28	g_adventure																				
29	g_comedy																				
30	g_drama																				
31	g_thriller																				
32	star_score																				
33	ensemble																				
34	lead_race																				
35	release_wide_0																				
36	release_wide_1																				
37	release_wide_2																				
38	re_release																				
39	max_theater																				
40	mon_no_1																				
41	mon_no_2																				
42	mon_no_3	**1.00**																			
43	mon_no_4	**-.09**	**1.00**																		
44	mon_no_5	**-.09**	**-.08**	**1.00**																	
45	mon_no_6	**-.10**	**-.09**	**-.09**	**1.00**																
46	mon_no_7	**-.08**	**-.07**	**-.07**	**-.08**	**1.00**															
47	mon_no_8	**-.11**	**-.09**	**-.09**	**-.10**	**-.09**	**1.00**														
48	mon_no_9	**-.09**	**-.08**	**-.08**	**-.09**	**-.08**	**-.10**	**1.00**													
49	mon_no_10	**-.10**	**-.09**	**-.09**	**-.10**	**-.08**	**-.11**	**-.09**	**1.00**												
50	mon_no_11	**-.10**	**-.09**	**-.09**	**-.10**	**-.09**	**-.11**	**-.09**	**-.10**	**1.00**											
51	mon_no_12	**-.11**	**-.09**	**-.09**	**-.10**	**-.09**	**-.11**	**-.10**	**-.11**	**-.11**	**1.00**										
52	year_2010	.00	-.01	-.03	.02	.02	.00	-.02	.00	.00	.00	**1.00**									
53	year_2011	.00	.03	-.01	-.03	.02	.00	.05	-.02	-.01	-.02	**-.11**	**1.00**								
54	year_2012	-.01	.00	.00	.01	-.05	-.01	.01	.00	.00	.01	**-.11**	**-.11**	**1.00**							
55	year_2013	.01	-.01	.00	-.01	.00	.02	.00	-.02	.01	.00	**-.11**	**-.11**	**-.11**	**1.00**						
56	year_2014	-.01	.01	.01	-.01	-.01	.01	-.01	.04	-.04	.00	**-.11**	**-.11**	**-.11**	**-.11**	**1.00**					
57	year_2015	.00	-.01	.01	-.02	.01	-.01	.00	.01	-.02	.02	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**1.00**				
58	year_2016	.00	.00	-.02	.00	**.05**	-.03	.01	-.02	.02	-.02	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**1.00**			
59	year_2017	-.01	-.01	.00	.02	-.01	-.03	.00	.02	-.01	.02	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**1.00**		
60	year_2018	.02	.00	-.02	.00	-.01	.01	.00	-.01	.02	.01	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**1.00**	
61	year_2019	-.01	.00	.05	.01	-.04	.03	-.03	.00	.03	-.02	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	**-.11**	1.00

Notes: Boldfaced values for correlations indicate significance at a 95% level.

Web Appendix B: Film Content Analysis Review

TABLE 1W. 3. Review of Select Literature Using Content Analysis for Films

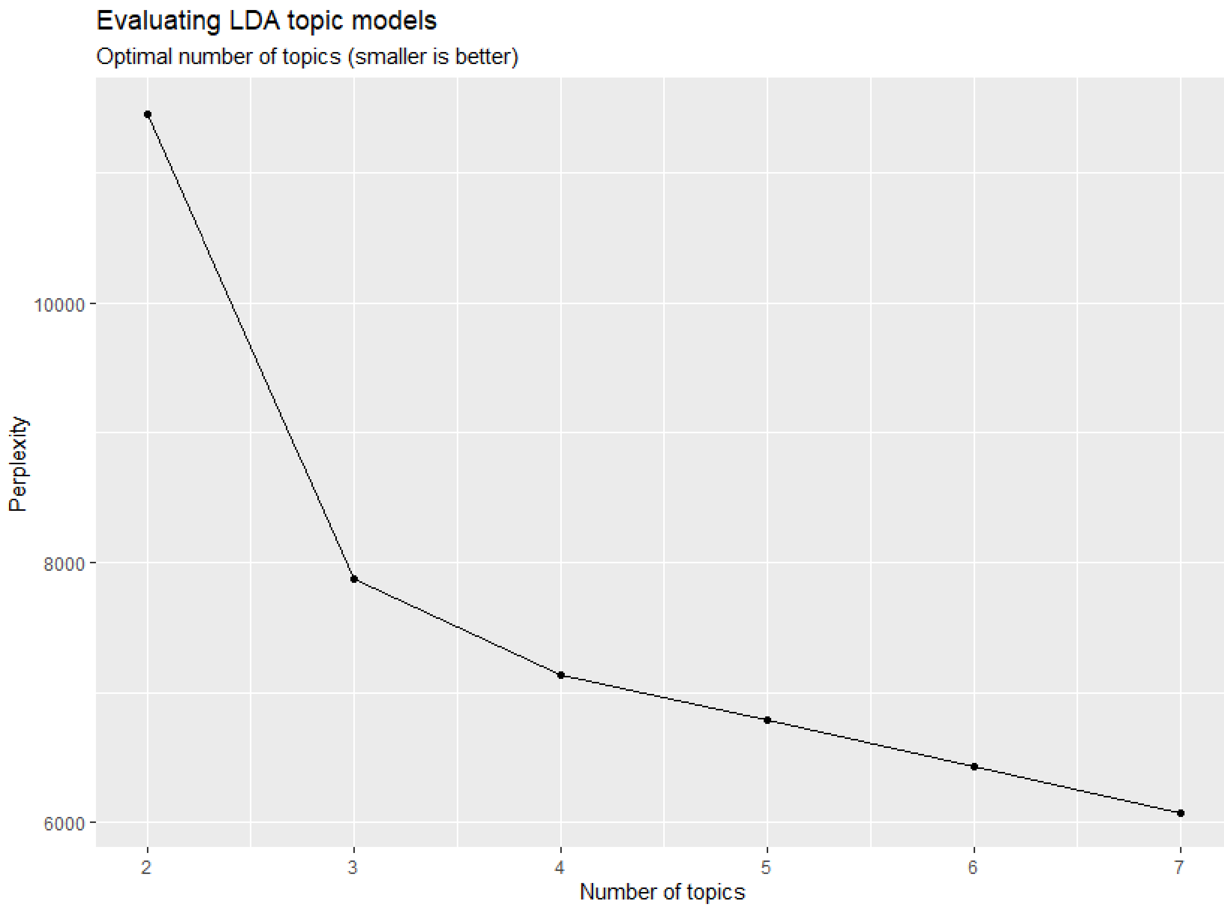
Work	Type of Content	Content perspective	Method	Purpose
(Kuppuswamy and Younkin 2020)	Wikipedia plot summary	Studio	LIWC	The authors used LIWC to categorize plot summary into story themes and used these themes as controls in the model
(Berger, Kim, and Meyer 2021)	Script	Studio	Dictionary-based	The author used pre-defined dictionaries to measure chunks of text in scripts in terms levels of emotional content and used the variance among the content as a measure of emotional volatility
(Paulich and Kumar 2021)	Script	Studio	LIWC	The authors used LIWC to measure emotional content
(Toubia et al. 2019)	IMDb plot synopsis	Studio	Independent judges	The authors used independent judges to code film content to validate an LDA algorithm with seed words
(Eliashberg, Hui, and Zhang 2007)	Movie spoilers	Studio	Independent raters based on a pre-defined questionnaire	The authors used content analysis of scripts to predict box office success
(Ryoo, Wang, and Lu 2021)	User reviews from IMDb	Audience review	CTM (extended LDA)	The authors used CTM to identify 61 topics from user reviews. These topics were used to identify spoiler vs. nonspoiler reviews and estimate spoiling intensity in reviews
(Xu et al. 2019)	Movie scripts and synopsis	Studio	Word embedding (dictionary-based)	The authors used word embedding around gendered characters to identify gender stereotypes in entertainment content
(Sreenivasan 2013)	IMDb keywords	Audience	Distribution of keywords over time	The authors used the distribution of keywords over time to assess the novelty of film content

Web Appendix C: Topic Modeling

Topic Selection

In this study, our main purpose of the topic modeling procedure was topic discovery, or uncovering themes (including stereotypes used) in films as seen through the consumer perspective (Humphreys and Wang 2018). Similar to Berger et al. (2021), we primarily sought to understand our current sample of films as opposed to predicting future contents. In determining the number of topics, we focused on balancing between statistical discrimination and parsimony. When the topics are not discriminant, documents (set of keywords) are distributed almost evenly across topics, defeating the purpose of understanding key themes in the texts. The higher the number of topics, the more likely that the topics are more discriminant from each other, yet the lower the interpretability of the topics (Berger et al. 2020). Following Berger et al. (2020), we combined the use of a perplexity statistic and careful evaluation of different topic solutions in topic selection. The perplexity statistic is a model-fit measure that estimates how well a probability model predicts a sample (Blei, Ng, and Jordan 2003). Lower perplexity indicates better model performance. We conducted multiple iterations of LDA with different numbers of topics (k = 2:7) and calculated the perplexity score of each solution (See Figure 1W.1). Because the perplexity statistic is equivalent to “the inverse of the geometric mean of the per-word likelihood” (Liu and Toubia 2018, p. 942), as the number of topics increases, the perplexity statistic always decreases, as shown in Figure 1W.1.

FIGURE 1W. 1. Topic Selection, Perplexity Measure

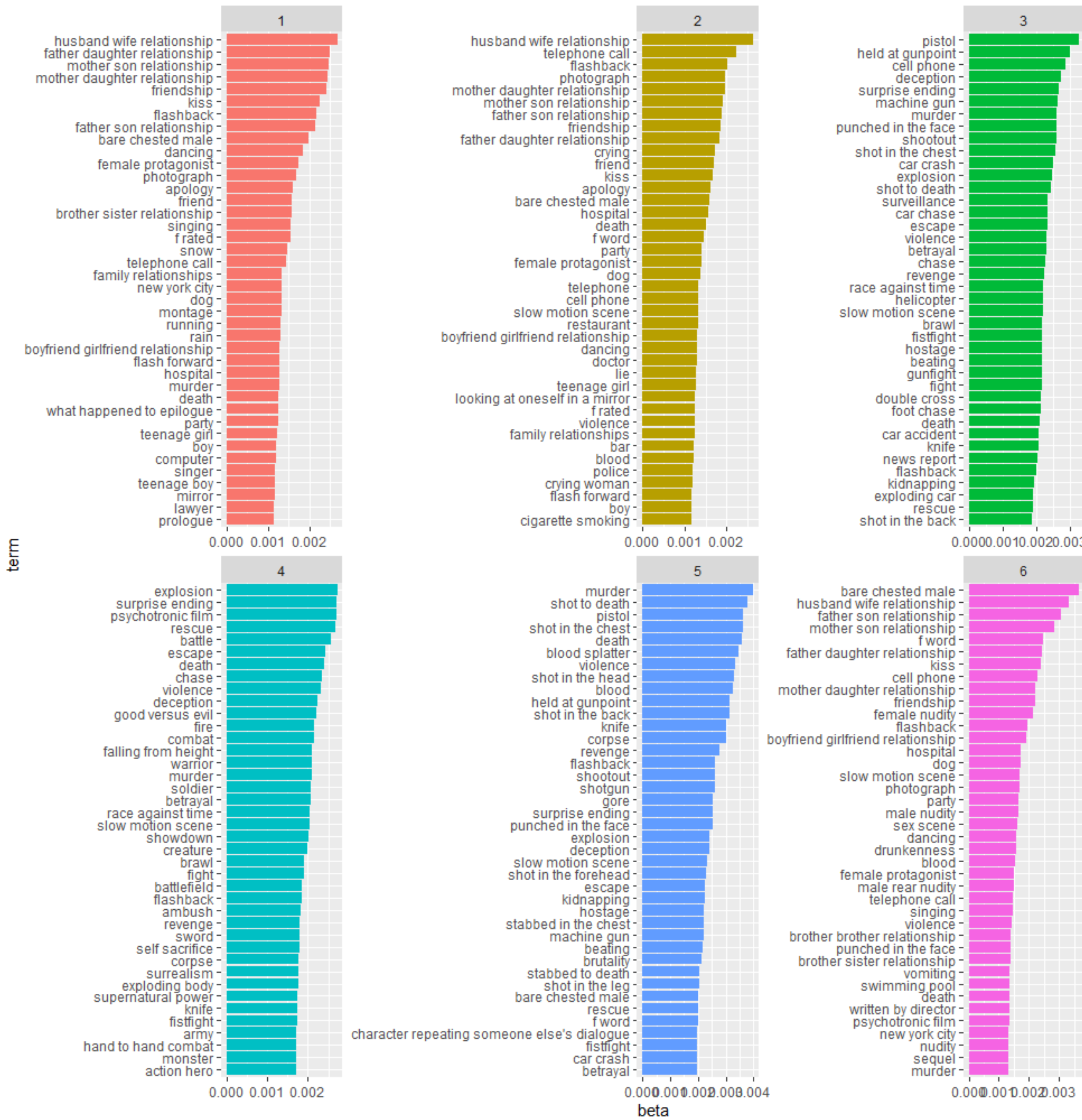


To balance between parsimony and perplexity, we carefully reviewed different solution results in terms of highest loaded keywords and most differing keywords among topics (calculated by log ratios); see Figure 1W.2 for examples of 2, 3, and 6 topics. We chose the final solution of three topics, as it represented the most drastic reduction in perplexity scores while maintaining parsimony.

FIGURE 1W. 2. Top 30 Highest Loading Keywords for k=2,3,6

NOTE.— Vertical axis: Keyword terms; Horizontal axis: probability per topic (beta)

Six-topic LDA



Note: Vertical axis: Keyword terms; Horizontal axis: probability per topic (beta)

Topic Labeling and Interpretation

To label these topics, we first examined the most likely keywords for each topic. As shown in Figure 1.3, we then studied the keywords that were most discriminant (that is, much more likely to appear in one topic as compared to another) to inform our categorization. We did so by using the likelihood estimated by the LDA procedure to calculate the log ratio of each keyword to belong to one topic relative to another topic[4]. For example, keyword "heroine saves hero" is most likely to represent topic 2 over topics 1 and 3 (as its log ratios over the other two topics are 42 and 435 respectively). Topic 1 is represented by keywords that signify violence such as "violence," "shot to death," "shot in the chest," "pistol," etc. We therefore labeled this topic *Brutal*. Topic 2, while also sharing violence-related keywords such as "explosion" or "death," prominently features keywords that focus on a positive theme such as "good versus evil," "friendship," "self-sacrifice," or a fantasy-oriented theme such as "supernatural," "fantasy world," or "fairy tale." We labeled this topic *Aspirational*. Finally, relationship-oriented keywords such as "family/ husband wife/ father son/ mother son/ mother daughter relationship" or social context keywords such as "telephone call, "kiss," "restaurant," and "hospital" have high loadings on topic 3; we labeled this topic "*Relational*."

We further reviewed the films with the highest loadings on each topic (see Table 4 for descriptions of topics, keywords, and example films). The lead characters in Topic *Brutal* were powerful and vicious and often alone (e.g., John Wick). Female characters in this topic fitted the damsel in distress archetype in films; they either needed protection (e.g., Lois Lane in *Men of*

[4] Log $ratio_{ij} = \log_2(\beta_i/\beta_j)$ in which β_i, β_i are the loadings of a word on topic i and j respectively. The log ratio statistic represents how big the difference between two topics is for a particular keyword. A log ratio of 0 means a word is equally likely to belong to both topics. A log ratio of 1 in topic 1 over topic 2 means that a word is twice more likely to belong to topic 1 than topic 2. This statistic is useful for identifying most differentiating keywords in one topic versus another.

Steel) or were victimized and then turned vengeful (e.g., *Lucy*). On the contrary, lead characters in Topic *Aspirational*, while similarly powerful, were other-focused, often in a quest for the common good (e.g., *Wonder Woman*, *Harry Potter*). Lead characters in Topic *Relational* were depicted in complex psychological struggles and vulnerabilities (e.g., *The King's Speech, A Star is Born*), counter-stereotypical for the masculine male archetypes in films.

We then conducted additional analysis of the topics focusing on the gendered keywords. We first ran all the unique keywords through Linguistic Inquiry and Word Count program (LIWC; Pennebaker et al. 2015) to identify gendered keywords and their valence. We used a series of binary variables to code keywords that featured positive (negative) female (male) terms. We then matched these keywords with their loadings on each topic and their relative log ratios between the topics. Keyword "woman held captive," a negative female keyword, for example, was more likely to be in Topic *Brutal* than Topics *Aspirational* and *Relational* (log ratios = 412 and 439 respectively). We then averaged the log ratios of these positive and negative gendered keywords for each topic (Table W4). The results show that on average, negative female keywords load much higher in topic *Brutal* than in Topic *Aspirational* (avg. log ratio = 96) while positive female keywords much more likely to appear in topic *Aspirational* than in topic *Brutal* (avg. log ratio = 9). Additionally, keywords that include "victim" load much higher in Topic *Brutal* than in Topic *Aspirational* (avg. log ratio =141). Both negative and positive female keywords are most likely to appear in Topic *Relational*, a context that stereotypically focuses on women. Interestingly, in this counter-stereotypical topic for male roles, positive male keywords (e.g., "father love") are also more prevalent in *Relational* than in the *Brutal* and *Aspirational* topics (avg. log ratios = 43 and 27 respectively).

TABLE 1W. 4. Average Log Ratios of Positive and Negative Gendered Keywords

Keywords	Example	Brutal / Aspirational	Brutal / Relational	Relational / Aspirational
Positive female (N=185)	"Superheroine," "female best friend," "mother hug"	-13.02	-61.11	48.07
Negative female (N=622)	"female killer," "bad mother," "woman held captive"	96.51	-8.66	105.17
Positive male (N=168)	"father love," "superhero," "good man"	-27.68	-43.55	15.87
Negative male (N=458)	"abusive father," "evil business man," "mad king"	66.57	4.93	61.64
Victim (N=37)	"female victim," "victim," "child abuse victim"	141.50	1.72	139.78

Note that while the gendered keywords were informative, they were not comprehensive. The keywords include only explicitly gendered terms, missing those that might refer to a certain gender but, because they fell into the gender stereotypes (e.g., "nurse," or "doctor"), the gender indicator was omitted. Keywords could also be both positive and negative at the same time (e.g., "tragic hero"), diluting our understanding. To further validate the topics, we followed up with a validation study using films in each topic with a general audience.

Web Appendix D: Topic Validation Study

This study aimed to validate the topics found in the topic modeling procedure. Here, we selected a set of four films for each topic (2 male-led, 2 female-led) and asked participants to evaluate the main female and male characters of each film in terms of positive (negative) stereotypical female (male) traits. We expected that participants perceived significant differences in gender portrayals across each topic, validating our topic modeling results.

Method

Participants, Design, Procedure, and Measures. We recruited 652 Amazon Mechanical Turk workers (M_{Age} = 37.07, SD_{Age}= 1.38, 35.99% female) to participate in a 3 (*topic*) × 2 (*lead gender*) between-subjects experiment. Under the guise of a movie evaluation task, each participant was randomly assigned to evaluate one film from a set of 12 pre-selected films. These 12 films (four per topic) were most representative of their respective topic (topic loadings of at least 98%) and balanced in terms of lead gender and release time (before or after #MeToo; for the full list of films used in this study, see Table 3). Each film was therefore evaluated approximately 50 times. Participants were first instructed to watch and review the film's official trailer and synopsis. Participants were then asked to complete an evaluation task of the film's main female and male characters (for detailed measures, see Table W5). Finally, participants completed an attention check and demographic questions. Thirty-eight participants failed the attention check ("Which movie did you review today?" out of three options) and were excluded from the study (final N = 614).

TABLE 1W. 5. Validation Study: Measures

Measures	Item
Gender portrayals	Based on the movie's trailer and synopsis, how much do you agree with the following statements regarding the main female (male) character of the film? *(1 = "Strongly Disagree," 7 = "Strongly Agree").* The main female (male) character is… *(Positive female stereotype item)* • …warm. • …other-focused. • …gentle. *(Negative female stereotype item)* • …weak. • …victimized. • …emotional. *(Positive male stereotype item)* • …powerful. • …competent. • …heroic. *(Negative male stereotype item)* • …cold. • …self-focused. • …violent.
Overall evaluation	Based on the movie's trailer and synopsis, how would you evaluate the main female (male) character of the film? • 1= "negative," 7 = "positive" • 1= "unfavorable," 7 = "favorable"
Film evaluation	Have you watched the film? (1 = "Yes," 0 = "No") • If yes: How much did you enjoy the film (1= "Not at all," 10= "Very much") • If no: How likely would you be willing to watch the film? (1= "Not at all," 10= "Very much")

Results

We averaged the items of each sub-category to derive scores for perceived positive (negative) female (male) stereotype traits of the lead characters in these films (αs >.70). We conducted a series of 3 (*topic*) × 2 (*lead gender*) factorial ANOVA on the derived measures, as well as each of the gender stereotype items.

Consistent with our hypothesis, the results revealed a significant main effect of topic on perceived positive female stereotypes and on positive and negative male stereotypes ($F(2, 608)$ = 9.39, 15.42, 7.38 respectively, ps <.001). Although the main effect of *topic* on perceived negative female stereotypes was not significant, sub-item analysis revealed a significant main effect of topic on "victim" ($F(2, 608) = 3.57$, $p = .03$) and a marginally significant main effect on "weak" ($F(2,608) = 2.81$, $p = .06$). Pairwise comparisons (see Table W6) showed that as compared to

Topics *Aspirational* and *Relational*, lead characters in Topic *Brutal* were rated significantly lower on positive stereotypical female traits and higher on negative stereotypical male traits. As compared to Topics *Aspirational* and *Brutal*, lead characters in Topic *Relational* were rated significantly lower in male stereotypes (both positive and negative).

The results showed a significant main effect of lead gender on female stereotypes, such that across films in our sample, male leads were seen as significantly higher in both positive and negative female stereotypes than female leads. We did not observe a significant interaction effect of *topic* × *lead* gender on female stereotypes (both positive and negative) and negative male stereotypes. We observed a significant interaction effect of *topic* × *lead gender* only on positive male stereotypes (F (2,608) = 3.48, $p = .03$), such that in Topic *Aspirational*, female (vs. male) lead characters were perceived as having more positive male stereotypes ("powerful," "competent," "heroic"), while such effect was reversed in Topic *Relational*. Table W6 summarizes the pairwise comparison of the topics across different measures.

TABLE 1W. 6. Validation Study Results

	Average Rating by Topic			**Pairwise Comparisons T-Tests**		
	Brutal (I)	**Aspirational (II)**	**Relational (III)**	**(I)-(II)**	**(I)-(III)**	**(II)-(III)**
Positive female stereotypes	4.88 (.09)	**5.37 (.08)**	5.28 (.09)	-.49 (.12) ***	-.41 (.12) ***	.09 (.12) *n.s.*
warm	4.99 (.10)	5.35 (.10)	**5.38 (.10)**	-.36 (.14) **	-.39 (.14) ***	-.03 (.14) *n.s.*
other-focused	5.03 (.11)	**5.44 (.10)**	5.25 (.11)	-.42 (.15) ***	-.22 (.15) *n.s.*	.19 (.15) *n.s.*
gentle	4.61 (.11)	**5.32 (.11)**	5.21 (.11)	-.71 (.16) ***	-.60 (.16) ***	.10 (.16) *n.s.*
Negative female stereotypes	4.70 (.10)	4.76 (.09)	**4.78 (.10)**	-.06 (.14) *n.s.*	-.08 (.14) *n.s.*	-.02 (.13) *n.s.*
victim	**5.04 (.12)**	4.71 (.12)	4.60 (.12)	.33 (.17) **	.44 (.17) **	.11 (.17) *n.s.*
weak	3.96 (.15)	4.30 (.14)	**4.45 (.15)**	-.34 (.21) *	-.49 (.21) **	-.15 (.20) *n.s.*
emotional	5.09 (.11)	5.27 (.10)	**5.28 (.11)**	-.18 (.15) *n.s.*	-.19 (.15) *n.s.*	-.01 (.15) *n.s.*
Positive male stereotypes	**5.71 (.08)**	5.53 (.07)	5.29 (.08)	.19 (.11) *	.42 (.11) ***	.23 (.11) **
powerful	**5.87 (.09)**	5.49 (.09)	5.31 (.09)	.38 (.13) ***	.56 (.13) ***	.17 (.13) *n.s.*
competent	**5.68 (.10)**	5.47 (.09)	5.32 (.09)	.21 (.13) *n.s.*	.36 (.13) ***	.15 (.13) *n.s.*
heroic	5.58 (.10)	**5.62 (.09)**	5.25 (.09)	-.04 (.13) *n.s.*	.33 (.14) **	.37 (.13) ***
Negative male stereotypes	**5.16 (.11)**	4.77 (.10)	4.35 (.10)	.40(.14) ***	.82 (.15) ***	.42 (.14) ***
violent	**5.37 (.13)**	4.77 (.12)	4.11 (.12)	.60 (.17) ***	1.27 (.18) ***	.66 (.17) ***
self-focused	**5.17 (.12**)	4.92 (.11)	4.60 (.12)	.25 (.16) *n.s.*	.57 (.17) ***	.32 (.16) *
cold	**4.95 (.13)**	4.60 (.12)	4.33 (.13)	.35 (.18) *	.62 (.19) ***	.28 (.18) *n.s.*

Note: Standard errors in parentheses. ***$p < .01$, **$p < .05$, *$p < .10$, *n.s*. not significant. **Bolded value** reflects the highest average rating per item among the three topics.

Finally, we did not find any significant main effect or interaction effect of topic and lead gender on the overall evaluation of the lead character or the film. In addition, we also tested the 3 (topic) × 2 (lead gender) × 2 (post: before or after #MeToo) interaction on all the measures and did not find any significant three-way interactions or main effect of *post*, ruling out the argument that films released after #MeToo might be fundamentally different in gender portrayals than those released before #MeToo.

Discussion

This study further validated the three topics generated from our topic modeling procedure. The results showed that films in Topic *Brutal* exhibited the most negative male and female stereotypical traits, characterizing the damsel-in-distress female and ultra-masculine male archetypes in films. In Topic *Aspirational*, female lead characters were powerful, heroic, and competent, counter-stereotypical for female characters. In Topic *Relational*, male lead characters were least likely to exhibit stereotypical male traits (both positive and negative ones). They were seen as the least violent, cold or powerful, and competent, but rather warm and gentle, counter-stereotypical for male characters in films.

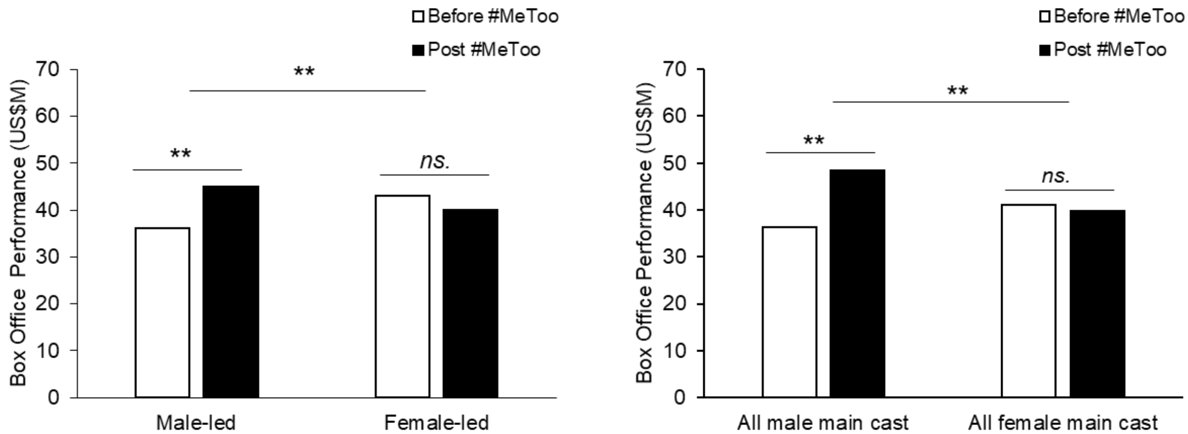
Web Appendix E: List Of Studios

TABLE 1W. 7. List of Major Studios and Number of Films Released from 2010 to 2019

Studio	**# films**
Lionsgate	166
Paramount	111
Sony	170
Twentieth Century Fox	159
Universal	211
Disney	74
Warner Bros	180
Others	255
Total	**1326**

Web Appendix F: Simple Slope Results

FIGURE 1W. 3. Main Effects of #MeToo on Box Office Performance of Female (vs. Male)-centric films



Note: ****p* < .01, ***p* < .05, **p* < .10, *ns.* not significant

TABLE 1W. 8. Simple Slopes of the #MeToo Effects across Different Operationalizations of Female Variable

	Coefficients of Post among female-centric films					Coefficients of Post among male-centric films				
	Coefficients	β	SE	t	p	Coefficients	β	SE	t	p
Topic Brutal	Post (female = f_lead)	-.26	.15	-1.72	.09 *	Post (male = m_lead)	.18	.14	1.29	.20
	Post (female = f_ratio)	-.59	.26	-2.24	.03 **	Post (male = m_ratio)	.41	.17	2.47	.01 **
	Post (female = f_rated)	-.49	.24	-2.01	.04 **	Post (male = m_rated)	.21	.14	1.54	.12
Topic Aspirational	Post (female = f_lead)	.12	.16	.72	.47	Post (male = m_lead)	.09	.14	.64	.53
	Post (female = f_ratio)	.01	.22	.03	.98	Post (male = m_ratio)	.21	.17	1.26	.21
	Post (female = f_rated)	.06	.21	.29	.77	Post (male = m_rated)	.11	.14	.79	.43
Topic Relational	Post (female = f_lead)	.08	.13	.59	.55	Post (male = m_lead)	.30	.12	2.44	.01 **
	Post (female = f_ratio)	.13	.16	.80	.43	Post (male = m_ratio)	.29	.15	1.98	.05 **
	Post (female = f_rated)	.18	.15	1.23	.22	Post (male = m_rated)	.21	.12	1.76	.08 *

Notes: ****p* < .01, ***p* < .05, **p* < .10

Web Appendix G: Additional Analyses

TABLE 1W. 9. Sensitivity Analysis with Different DVs

	Studio fixed-effects model							
	DV= Global Box Office Performance (log transformed)				DV= First Month US Box Office (log transformed)			
Variables	Model 1 Female=f_lead		Model 2 Female = f_ratio		Model 3 Female=f_lead		Model 4 Female = f_ratio	
N = 1312	β	(SE)	β	(SE)	β	(SE)	β	(SE)
Female × Post × topicAspirational	.47	(.27)	.29	(.53)	.43 **	(.21)	.60	(.42)
Female × Post × topicRelational	.10	(.24)	.72	(.46)	.13	(.19)	.77 **	(.37)
Female × Post	-.30	(.21)	-.81 **	(.40)	-.33 **	(.16)	-.97 ***	(.31)
Post × topicAspirational	-.16	(.14)	-.09	(.22)	-.05	(.11)	-.09	(.17)
Post × topicRelational	.10	(.12)	-.13	(.19)	.15	(.10)	-.08	(.15)
Female × topicAspirational	.01	(.13)	.10	(.27)	.09	(.10)	-.21	(.22)
Female × topicRelational	.13	(.11)	.09	(.23)	.05	(.09)	-.02	(.18)
Post	.23	(.17)	.42 **	(.20)	.12	(.13)	.36 **	(.16)
Female	-.00	(.10)	.14	(.20)	.04	(.08)	.16	(.15)
Topic=Aspirational	.06	(.07)	.02	(.10)	-.00	(.05)	.08	(.08)
Topic=Relational	.11	(.07)	.11	(.10)	.12 **	(.05)	.11	(.08)
log(adj_bud)	.21 ***	(.02)	.21 ***	(.02)	.05 ***	(.02)	.05 ***	(.02)
meta_score	.01 ***	(.00)	.01 ***	(.00)	.01 ***	(.00)	.01 ***	(.00)
imax_3d	.16 **	(.07)	.16 **	(.07)	.06	(.06)	.06	(.06)
adapt	.02	(.04)	.02	(.04)	-.00	(.03)	-.00	(.03)
Sequel=1	.58 ***	(.08)	.58 ***	(.08)	.45 ***	(.06)	.45 ***	(.06)
Sequel=2	.41 ***	(.06)	.41 ***	(.06)	.22 ***	(.04)	.24 ***	(.04)
mpaaPG-13	-.58 **	(.23)	-.65 ***	(.23)	-.83 ***	(.19)	-.94 ***	(.19)
mpaaPG	-.45 *	(.23)	-.51 **	(.23)	-.65 ***	(.18)	-.76 ***	(.18)
mpaaR-Rated	-.54 **	(.23)	-.61 ***	(.23)	-.68 ***	(.18)	-.80 ***	(.18)
g_action	-.10 *	(.06)	-.10 *	(.06)	-.07	(.05)	-.09	(.05)
g_adventure	.16 ***	(.06)	.16 ***	(.06)	.00	(.04)	.00	(.04)
g_comedy	-.21 ***	(.04)	-.21 ***	(.04)	-.10 ***	(.03)	-.10 ***	(.04)
g_drama	-.09 *	(.05)	-.10 **	(.05)	-.05	(.04)	-.05	(.04)
g_thriller	.09	(.05)	.08	(.05)	-.01	(.04)	-.00	(.04)
star_score	.00 ***	(.00)	.00 ***	(.00)	.00 ***	(.00)	.00 ***	(.00)
ensemble	.24 ***	(.07)	.24 ***	(.07)	.20 ***	(.05)	.20 ***	(.05)
lead_race	.02	(.04)	.01	(.04)	.16 ***	(.04)	.16 ***	(.04)
release_wide=1	-.61 ***	(.07)	-.59 ***	(.07)	.42 ***	(.06)	.43 ***	(.06)
release_wide=2	.19 **	(.09)	.20 **	(.09)	-.42 ***	(.07)	-.43 ***	(.07)
re_release	1.58 ***	(.32)	1.55 ***	(.32)	.12	(.25)	.11	(.25)
max_theater	.00 ***	(.00)	.00 ***	(.00)	.00 ***	(.00)	.00 ***	(.00)
R^2	.73		.73		.76		.76	

Note: Standard errors in parentheses. ****p* < .01, ***p* < .05, **p* < .10

TABLE 1W. 10. Sensitivity Analysis among Films that Load at Least 50% in One Topic

Studio fixed-effects model				
	DV: US Box Office Performance (log transformed)			
	Model 1		Model 2	
Variables	Female=f_lead		Female = f_ratio	
N = 1298	β	(SE)		(SE)
Female × Post × topicAspirational	.21	(.20)	.86 **	(.38)
Female × Post × topicRelational	.47 **	(.22)	.79	(.44)
Female × Post	-.44 ***	(.17)	-1.01 ***	(.33)
Post × topicAspirational	.13	(.10)	-.13	(.16)
Post × topicRelational	-.09	(.12)	-.20	(.18)
Female × topicAspirational	.05	(.09)	-.04	(.19)
Female × topicRelational	.09	(.11)	-.22	(.23)
Post	.19	(.14)	.42 **	(.17)
Female	.07	(.08)	.19	(.16)
Topic=Aspirational	.12 **	(.06)	.12	(.08)
Topic=Relational	-.03	(.05)	.05	(.09)
log(adj_bud)	.05 **	(.02)	.05 **	(.02)
meta_score	.01 ***	(.00)	.01 ***	(.00)
imax_3d	.03	(.06)	.03	(.06)
adapt	.00	(.03)	.01	(.03)
Sequel=1	.57 ***	(.06)	.57 ***	(.06)
Sequel=2	.26 ***	(.05)	.27 ***	(.05)
mpaaPG-13	-.17	(.19)	-.28	(.19)
mpaaPG	-.10	(.19)	-.21	(.19)
mpaaR-Rated	-.20	(.19)	-.32	(.19)
g_action	-.12 **	(.05)	-.14 ***	(.05)
g_adventure	.00	(.05)	.00	(.05)
g_comedy	-.10 ***	(.04)	-.10 ***	(.04)
g_drama	-.03	(.04)	-.03	(.04)
g_thriller	-.02	(.04)	-.02	(.04)
star_score	.00 ***	(.00)	.00 ***	(.00)
ensemble	.22 ***	(.06)	.22 ***	(.06)
lead_race	.16 ***	(.04)	.15 ***	(.04)
release_wide=1	-.32 ***	(.06)	-.31 ***	(.06)
release_wide=2	-.30 ***	(.08)	-.31 ***	(.08)
re_release	-.18	(.30)	-.21	(.30)
max_theater	.00 ***	(.00)	.00 ***	(.00)
R^2	.74		.74	

Note: Standard errors in parentheses. ****p* < .01, ***p* < .05, **p* < .10

TABLE 1W. 11. Robustness Check: Different Operationalizations of Female Variable

Studio fixed-effects model				
	DV: US Box Office Performance (log transformed)			
Variables	Model 1 (N=1157) Female=Bechdel		Model 2 (N=1312) Female = f_rated	
	β	(SE)	β	(SE)
Female × Post × topicAspirational	.06	(.21)	.65 **	(.29)
Female × Post × topicRelational	.05	(.18)	.67 ***	(.25)
Female × Post	-.19	(.14)	-.70 ***	(.22)
Post × topicAspirational	.04	(.15)	-.10	(.11)
Post × topicRelational	.10	(.13)	.00	(.09)
Female × topicAspirational	-.03	(.09)	-.02	(.20)
Female × topicRelational	.01	(.08)	-.17	(.18)
Post	.37 **	(.17)	.21	(.14)
Female	.08	(.06)	.13	(.17)
Topic=Aspirational	-.01	(.07)	-.02	(.05)
Topic=Relational	.14 **	(.07)	.15 ***	(.05)
log(adj_bud)	.05 **	(.02)	.04 **	(.02)
meta_score	.01 ***	(.00)	.01 ***	(.00)
imax_3d	.03	(.06)	.04	(.06)
adapt	.01	(.03)	.01	(.03)
Sequel=1	.53 ***	(.06)	.56 ***	(.06)
Sequel=2	.23 ***	(.05)	.27 ***	(.05)
mpaaPG-13	-.16	(.24)	-.21	(.19)
mpaaPG	-.07	(.23)	-.13	(.19)
mpaaR-Rated	-.18	(.23)	-.25	(.19)
g_action	-.15 ***	(.05)	-.14 ***	(.05)
g_adventure	.02	(.05)	.01	(.05)
g_comedy	-.10 **	(.04)	-.10 ***	(.04)
g_drama	-.04	(.04)	-.03	(.04)
g_thriller	-.02	(.05)	-.02	(.04)
star_score	.00 ***	(.00)	.00 ***	(.00)
ensemble	.21 ***	(.06)	.22 ***	(.06)
lead_race	.13 ***	(.04)	.15 ***	(.04)
release_wide=1	-.32 ***	(.06)	-.31 ***	(.06)
release_wide=2	-.31 ***	(.08)	-.30 ***	(.08)
re_release	-.24	(.26)	-.21	(.26)
max_theater	.00 ***	(.00)	.00 ***	(.00)
R^2	.74		.74	

Note: Standard errors in parentheses. ****p* < .01, ***p* < .05, **p* < .10

References

Adler, Freda (1975), *Sisters in Crime: The Rise of The New Female Criminal.*, McGraw-Hill.

Ahmed, Sumaiya and Ashish Sinha (2016), "When It Pays to Wait: Optimizing Release Timing Decisions for Secondary Channels in the Film Industry," *Journal of Marketing*, 80 (4), 20–38.

Amenta, Edwin, Neal Caren, Elizabeth Chiarello, and Yang Su (2010), "The Political Consequences of Social Movements," *Annual Review of Sociology*, 36, 287–307.

Amenta, Edwin and Francesca Polletta (2019), "The Cultural Impacts of Social Movements," *Annual Review of Sociology*, 45, 279–99.

Anderson, Monica, Michael Barthel, Andrew Perrin, and Emily A. Vogels (2020), "#BlackLivesMatter Hashtag Surges on Twitter after George Floyd's Death | Pew Research Center."

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996), "Identification of Causal Effects using Instrumental Variables," *Journal of the American Statistical Association*, 91 (434), 444–55.

Atwater, Leanne E, Rachel E Sturm, Scott N Taylor, and Allison Tringale (2021), "The Era of #MeToo and What Managers Should Do about It," *Business Horizons*, 64 (2), 307–18.

Ballhaus, Werner and Wilson Chow (2021), "Perspectives from the Global Entertainment & Media Outlook 2021-2025 |PwC."

Basuroy, Suman and Subimal Chatterjee (2008), "Fast and Frequent: Investigating Box Office Revenues of Motion Picture Sequels," *Journal of Business Research*, 61 (7), 798–803.

Benford, Robert D (1993), "'You Could be the Hundredth Monkey': Collective Action Frames and Vocabularies of Motive within the Nuclear Disarmament Movement," *Sociological Quarterly*, 34 (2), 195–216.

Benford, Robert D and David A Snow (2000), "Framing Processes and Social Movements: An Overview and Assessment," *Annual Review of Sociology*, 26 (1), 611–39.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W Moe, Oded Netzer, and David A Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Berger, Jonah, Yoon Duk Kim, and Robert Meyer (2021), "What Makes Content Engaging? How Emotional Dynamics Shape Success," *Journal of Consumer Research*.

Bernstein, Paula (2015), "Sundance 2015 Infographic: Most Festival Films Will Land Distribution Deals | IndieWire," *IndieWire*.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan (2015), "Gender Identity and Relative Income within Households," *The Quarterly Journal of Economics*, 130 (2), 571–614.

Bhagwat, Yashoda, Nooshin L Warren, Joshua T Beck, and George F Watson IV (2020), "Corporate Sociopolitical Activism and Firm Value," *Journal of Marketing*, 84 (5), 1–21.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Box Office Mojo (2021), "Domestic Yearly Box Office - Box Office Mojo," *Box Office Mojo*, (accessed January 8, 2022), [available at https://www.boxofficemojo.com/year/?grossesOption=totalGrosses].

Buchanan, Larry, QuocTrung Bui, and Jugal K. Patel (2020), "Black Lives Matter May Be the Largest Movement in U.S. History - The New York Times," *The New York Times*, (accessed November 22, 2021), [available at https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html].

Carlsen, Audrey, Salam Maya, claire cain Miller, Denise Lu, Ash Ngu, Jugal K. Patel, and Zach Wichter (2018), "#MeToo Brought Down 201 Powerful Men. Nearly Half of Their Replacements Are Women. - The New York Times," *The New York Times*.

Chow, Andrew (2020), "How Parasite's Historic Oscars Night Happened | Time," *TIME*.

Davies, Michelle and Paul Rogers (2006), "Perceptions of Male Victims in Depicted Sexual Assaults: A Review of the Literature," *Aggression And Violent Behavior*, 11 (4), 367–77.

DeFronzo, James and Jungyun Gill (2019), *Social Problems and Social Movements*, Rowman & Littlefield.

Diani, Mario (1992), "The Concept of Social Movement," *The Sociological Review*, 40 (1), 1–25.

Eagly, Alice H and Steven J Karau (2002), "Role Congruity Theory of Prejudice toward Female Leaders," *Psychological Review*, 109 (3), 573.

Eagly, Alice H and Antonio Mladinic (1989), "Gender Stereotypes and Attitudes toward Women and Men," *Personality and Social Psychology Bulletin*, 15 (4), 543–58.

Einav, Liran (2007), "Seasonality in the US Motion Picture Industry," *The Rand Journal of Economics*, 38 (1), 127–45.

Elberse, Anita (2008), “The Power of Stars: Do Star Actors Drive the Success of Movies?,” *Journal of Marketing*, 71 (4), 102–20.

Eliashberg, Jehoshua, Anita Elberse, and Mark A A M Leenders (2006), “The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions,” *Marketing Science*, 25 (6), 638–61.

Eliashberg, Jehoshua, Sam K Hui, and Z John Zhang (2007), “From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts,” *Management Science*, 53 (6), 881–93.

Ellemers, Naomi (2018), “Gender Stereotypes,” *Annual Review of Psychology*, 69, 275–98.

Erigha, Maryann (2015), “Race, Gender, Hollywood: Representation in Cultural Production and Digital Media’s Potential for Change,” *Sociology Compass*, 9 (1), 78–89.

Ezzedeen, Souha R (2015), “Portrayals of Career Momen in Hollywood films: Implications for the Glass Ceiling’s Persistence,” *Gender in Management: An International Journal*.

Flood, Michael (2019), “Men and# MeToo: Mapping Men’s Responses to Anti-Violence Advocacy,” in *# MeToo and The Politics of Social Change*, Springer, 285–300.

Follows, Stephen (2018), “How Long does the Average Hollywood Movie Take to Make?”

Grau, Stacy Landreth and Yorgos C Zotos (2016), “Gender Stereotypes in Advertising: A Review of Current Research,” *International Journal of Advertising*, 35 (5), 761–70.

Guinier, Lani and Gerald Torres (2013), “Changing the Wind: Notes toward a Demosprudence of Law and Social Movements,” *Yale LJ*, 123, 2740.

Heath, Timothy B, Subimal Chatterjee, Suman Basuroy, Thorsten Hennig-Thurau, and Bruno Kocher (2015), “Innovation Sequences over Iterated Offerings: A Relative Innovation, Comfort, and Stimulation Framework of Consumer Responses,” *Journal of Marketing*, 79 (6), 71–93.

Hilton, James L and William Von Hippel (1996), “Stereotypes,” *Annual Review of Psychology*, 47 (1), 237–71.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), “Automated Text Analysis for Consumer Research,” *Journal of Consumer Research*, 44 (6), 1274–1306.

IMDb (2021), “IMDb | Help,” *The Internet Movie Database*, (accessed January 6, 2022), [available at https://help.imdb.com/article/contribution/titles/keywords/GXQ22G5Y72TH8MJ5#].

Janiszewski, Chris and Robert S Wyer Jr (2014), “Content and Process Priming: A Review,” *Journal of Consumer Psychology*, 24 (1), 96–118.

Jasper, James M (2011), “Emotions and Social Movements: Twenty Years of Theory and Research,” *Annual Review of Sociology*, 37, 285–303.

Joshi, Amit and Huifang Mao (2012), “Adapting to Succeed? Leveraging the Brand Equity of Best Sellers to Succeed at the Box Office,” *Journal of the Academy of Marketing Science*, 40 (4), 558–71.

Jost, John T and Aaron C Kay (2005), “Exposure to Benevolent Sexism and Complementary Gender Stereotypes: Consequences for Specific and Diffuse Forms of System Justification.,” *Journal of Personality and Social Psychology*, 88 (3), 498.

Judd, Charles M and Bernadette Park (1993), “Definition and assessment of accuracy in social stereotypes.,” *Psychological review*, 100 (1), 109.

King, Brayden G and Nicholas A Pearce (2010), “The Contentiousness of Markets: Politics, Social movements, and Institutional Change in Markets,” *Annual Review of Sociology*, 36, 249–67.

Klein, Jill Gabrielle, N Craig Smith, and Andrew John (2004), “Why We Boycott: Consumer Motivations for Boycott Participation,” *Journal of Marketing*, 68 (3), 92–109.

Kotler, Philip and Christian Sarkar (2018), *Brand Activism: From Purpose to Action*, IDEA BITE PRESS.

Kozinets, Robert V and Jay M Handelman (2004), “Adversaries of Consumption: Consumer Movements, Activism, and Ideology,” *Journal of Consumer Research*, 31 (3), 691–704.

Kroll, Justin (2021), “‘West Side Story’s Rachel Zegler To Play Disney’s Snow White In Movie – Deadline,” *Deadline*.

Kuppuswamy, Venkat and Peter Younkin (2020), “Testing the Theory of Consumer Discrimination as an Explanation for the Lack of Minority Hiring in Hollywood Films,” *Management Science*, 66 (3), 1227–47.

Linz, Daniel G, Edward Donnerstein, and Steven Penrod (1988), “Effects of Long-term Exposure to Violent and Sexually Degrading Depictions of Women.,” *Journal of Personality and Social Psychology*, 55 (5), 758.

Liu, Jia and Olivier Toubia (2018), “A semantic approach for estimating consumer content preferences from online search queries,” *Marketing Science*, 37 (6), 930–52.

Lopes, Amandha Rohr (2014), “The Impact of Social Media On Social Movements: The New Opportunity and Mobilizing Structure,” *Journal of Political Science Research*, 4 (1), 1–23.

Luo, Hong and Laurina Zhang (2021), “Scandal, Social Movement, and Change: Evidence from# Metoo in Hollywood,” *Management Science*, 1–19.

McCarthy, John D and Mayer N Zald (1977), "Resource Mobilization and Social Movements: A Partial Theory," *American Journal of Sociology,* 82 (6), 1212–41.

McDonald, Aubri F (2019), "Framing# MeToo: Assessing the Power and Unintended Consequences of a Social Media Movement to Address Sexual Assault," in *Handbook of Sexual Assault and Sexual Assault Prevention,* Springer, 79–107.

McDonnell, Mary-Hunter and Brayden King (2013), "Keeping Up Appearances: Reputational Threat and Impression Management after Social Movement Boycotts," *Administrative Science Quarterly,* 58 (3), 387–419.

"MeToo Movement" (2021), (accessed February 6, 2022), [available at https://metoomvmt.org/].

Miller, Monica K, Jessica Rauch, and Tatyana Kaplan (2016), "Gender differences in movie superheroes' roles, appearances, and violence," *A Journal of Gender New Media & Technology*.

Moon, Sangkil, Paul K Bergey, and Dawn Iacobucci (2010), "Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction," *Journal of Marketing,* 74 (1), 108–21.

Morgan, Stephen L and David J Harding (2006), "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice," *Sociological Methods & Research,* 35 (1), 3–60.

Morris, Aldon (2000), "Reflections on Social Movement Theory: Criticisms and Proposals," *Contemporary Sociology,* 29 (3), 445–54.

Morris, Amanda (2018), "#HimToo: Left And Right Embrace Opposing Takes On Same Hashtag : NPR," *NPR*.

Mukherjee, Anirban and Vrinda Kadiyali (2011), "Modeling Multichannel Home Video Demand in The Us Motion Picture Industry," *Journal of Marketing Research,* 48 (6), 985–95.

N'Duka, Amanda (2019), "Halle Bailey Set To Play Ariel In 'The Little Mermaid' Remake – Deadline," *Deadline*.

Nardini, Gia, Tracy Rank-Christman, Melissa G Bublitz, Samantha N N Cross, and Laura A Peracchio (2021), "Together We Rise: How Social Movements Succeed," *Journal of Consumer Psychology,* 31 (1), 112–45.

Paulich, Brianna JeeWon and V Kumar (2021), "Relating Entertainment Features in Screenplays to Movie Performance: An Empirical Investigation," *Journal of the Academy of Marketing Science,* 1–21.

Pennebaker, James W, Ryan L Boyd, Kayla Jordan, and Kate Blackburn (2015), "The development and psychometric properties of LIWC2015."

Petersen, Mitchell A (2009), “Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches,” *The Review of Financial Studies*, 22 (1), 435–80.

PettyJohn, Morgan E, Finneran K Muzzey, Megan K Maas, and Heather L McCauley (2019), “# HowIWillChange: Engaging Men and Boys in the# MeToo Movement.,” *Psychology of Men & Masculinities*, 20 (4), 612.

Pew Research Center (2018), “How Social Media Users Have Discussed Sexual Harassment Since #MeToo Went Viral.”

Della Porta, Donatella and Mario Diani (2020), *Social Movements: An Introduction*, John Wiley & Sons.

Pouliot, Louise and Paul S Cowen (2007), “Does Perceived Realism Really Matter in Media Effects?,” *Media Psychology*, 9 (2), 241–59.

Reinecke, Juliane and Shahzad Ansari (2021), “Microfoundations of Framing: The Interactional Production of Collective Action Frames in the Occupy Movement,” *Academy of Management Journal*, 64 (2), 378–408.

Rudman, Laurie A and Peter Glick (2001), “Prescriptive Gender Stereotypes and Backlash toward Agentic Women,” *Journal of Social Issues*, 57 (4), 743–62.

Russell, Helen (2021), “Lego to Remove Gender Bias from its Toys after Findings of Child Survey,” *The Guardian*.

Ryoo, Jun Hyun, Xin Wang, and Shijie Lu (2021), “Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue,” *Journal of Marketing*, 85 (2), 70–88.

Sen, Sankar, Zeynep Gürhan-Canli, and Vicki Morwitz (2001), “Withholding Consumption: A Social Dilemma Perspective on Consumer Boycotts,” *Journal of Consumer research*, 28 (3), 399–417.

Smith, Stephen M and Richard E Petty (1996), “Message framing and persuasion: A message processing analysis,” *Personality and Social Psychology Bulletin*, 22 (3), 257–68.

Sreenivasan, Sameet (2013), “Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords,” *Scientific reports*, 3 (1), 1–11.

Starkey, Jesse C, Amy Koerber, Miglena Sternadori, and Bethany Pitchford (2019), “#Metoo Goes Global: Media Framing of Silence Breakers in Four National Settings,” *Journal of Communication Inquiry*, 43 (4), 437–61.

Stoodard, Thomas B (1997), “Bleeding Heart: Reflections on Using the Law to Make Social Change,” *NYUL Rev.*, 72, 967.

"The Bechdel Test" (2019), *The Bechdel Test,* [available at https://bechdeltest.com/].

Toubia, Olivier, Garud Iyengar, Renée Bunnell, and Alain Lemaire (2019), "Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption," *Journal of Marketing Research,* 56 (1), 18–36.

Turner, Ralph H (1969), "The Theme of Contemporary Social Movements," *The British Journal of Sociology,* 20 (4), 390–405.

Tversky, Amos and Daniel Kahneman (1981), "The Framing of Decisions and the Psychology Of Choice," *Science,* 211 (4481), 453–58.

Valdivia, Angharad N (1998), "Clueless in Hollywood: Single Moms in Contemporary Family Movies," *Journal of Communication Inquiry,* 22 (3), 272–92.

Williams, John E and Deborah L Best (1990), *Measuring Sex Stereotypes: A Multination Study, Rev,* Sage Publications, Inc.

Xiong, Ying, Moonhee Cho, and Brandon Boatwright (2019), "Hashtag Activism and Message Frames among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter During the# Metoo Movement," *Public Relations Review,* 45 (1), 10–23.

Xu, Huimin, Zhang Zhang, Lingfei Wu, and Cheng-Jun Wang (2019), "The Cinderella Complex: Word Embeddings Reveal Gender Stereotypes in Movies and Books," *PloS One,* 14 (11), e0225385.

Yeung, Amy W Y, Aaron C Kay, and Jennifer M Peach (2014), "Anti-Feminist Backlash: The Role of System Justification in the Rejection Of Feminism," *Group Processes & Intergroup Relations,* 17 (4), 474–84.

This page is intentionally left blank

ESSAY II:
FEELING VIOLATED: HOW AND WHEN PRIVACY VIOLATIONS PRODUCE COMMENSURATE CONSUMER RESPONSES

In April 2019, Facebook's CEO Mark Zuckerberg declared to the world that "The future is private" and that the world's largest social media platform would be pivoting to a privacy-focused strategy (Nivea 2019). This denotes a significant shift from his own declaration in 2010 that privacy was an obsolete social norm of the past (Kirkpatrick 2010). Other companies have followed suit, changing their regard for privacy from a burdensome cost for risk management (Bamberger and Mulligan 2010) to a growing recognition of privacy as a potential source of competitive advantage. Apple, for example, recently released a major ad campaign with the simple tagline: "Privacy. That's iPhone." (Wuerthele 2019). Samsung has followed suit with its own privacy-focused campaign (Friedman 2022). Similarly, Google published a statement detailing its tech-driven approaches to enhancing consumer privacy protections while maintaining advertising personalization (Biron 2019).

Clearly privacy is more important to consumers than ever before (Kim, Barasz, and John 2019; Krishna 2020), due to growing digitization of consumption (Schmitt 2019) combined with marketers' exploitation of consumer data over the past decade (Benes 2018). More surprisingly, while consumers report unprecedented levels of privacy concerns (PwC 2017) and lack of control over their personal information (Pew 2019), they continue to patronize the products and services of many of the companies that routinely violate their privacy. For instance, consumers tolerate intrusive privacy practices by companies such as Amazon (Paul 2020), while similarly intrusive privacy violations by Facebook (Shane 2018) and Google (O'Flaherty 2018) have shifted substantial demand away from these brands and toward competitors (Mathews 2021).

For all its importance to modern consumption, the explanation for consumers' seemingly paradoxical privacy behaviors remain poorly understood (Acquisti, Brandimarte, and

Loewenstein 2020; Krasnova et al. 2010). This is exacerbated in part by a lack of clarity around the meaning of privacy (Acquisti, Taylor, and Wagman 2016; Appel et al. 2020; Martin and Murphy 2017) and privacy violation (MacInnis et al. 2020) as consumer constructs. Therefore, the main objective and contribution of the current research is to reconcile seemingly contradictory consumer responses to differing privacy situations. Toward this end, we theorize, test, and find converging support for a conceptual model, based on theories of control, that is both parsimonious and empirically robust.

As a springboard for developing this model, we also provide a theoretically-derived and empirically-confirmed conceptualization of consumer privacy violation. Specifically, the scope of our investigation examines privacy violations as consisting of three ways in which companies may use consumers' personal information once it has already been collected, reflecting a progressive increase in perceived severity: *recording* (i.e., merely observing and storing consumers' personal information), *targeting* (i.e., using consumers' information for targeted advertisements), and *sharing* (i.e., sharing consumers' information with a third party). Our conceptualization does not include unsanctioned data collection or breeches, but rather focuses on the ways that companies use the personal information that consumers choose to divulge. This reflects commonplace consumer experience, where consumers surrender personal data to companies but are typically uncertain as to how the data will be used, thus reflecting a violation in the sense that it detracts from consumers' sense of control (Puntoni et al. 2021; Walker 2016).

More specifically, through synthesizing literatures across domains of consumer privacy (Martin and Murphy 2017; Smith, Dinev, and Xu 2011), control (Deci and Ryan 2000; Weisz and Stipek 1982) and coping (Dweck 2000), we theorize that more severe privacy violations (e.g., *sharing* vs. *targeting*) reduce consumers' resource control (i.e., their subjective control over a valued resource; Fast et al. 2009). We predict that consumers will cope with this loss of control differently depending on the level of contingency provided by the marketplace. When market conditions provide a relatively high level of contingency (i.e., consumption outcomes are contingent upon a consumer's own behavior; Weisz 1980; Weisz and Stipek 1982), regaining resource control seems feasible, and consumers will therefore engage in active coping (e.g., through brand switching; Schiele and Venkatesh 2016). Conversely, when market conditions undermine contingency, regaining resource control is perceived as impossible and the subsequent sense of low contingency inhibits such control-reclaiming behaviors (Weisz 1980) such that consumers cope more passively (Dweck 2000).

In a market setting, this crucial boundary condition may manifest in several ways. For example, the variability of privacy practices in a given industry (henceforth "privacy variability;" tested in studies 1 and 4) may dictate the level of consumers' perceived contingency in that industry. Consequently, we expect a severe (vs. moderate) privacy violation to prompt greater control-reclaiming behaviors, only when the violating company operates in an industry with high privacy variability. This is because, in this market context, a desired outcome of consumption (e.g., increased control over one's own information) is perceived to be contingent upon a control-reclaiming behavior (e.g., brand switching). However, in an industry with lower privacy variability (e.g., in which companies share a standardized privacy practice), consumers are unlikely to engage in control-reclaiming behaviors regardless of violation severity because consumption outcomes seem non-contingent upon consumers' own actions. Similarly, intensity of competition in an industry (tested in studies 2 and 3) may serve as a manifestation of contingency, whereby industries with high (low) competitiveness are associated with high (low) contingency. If correct, our theorizing parsimoniously explains the phenomenon of consumers' sometimes failing to act in response to severe privacy violations, as exemplified by consumers' backlash against Facebook or Google but complacency with Amazon.

The remainder of this manuscript proceeds as follows. First, we present a rigorous review of several literatures spanning consumer privacy, control, and coping, leading to predictions around the meaning of privacy violation and when it translates into consumer action. We then use a multi-method approach to test this conceptual model across five studies. Using a between-participants experiment, study 1 first tests an industry-level manifestation of contingency—privacy variability—that helps explain opposing consumer responses to otherwise equivalent privacy violations across companies. Study 2 then replicates these findings in a new context using another industry-level characteristic to operationalize contingency, intensity of competition. Study 3 further replicates the moderating role of competitiveness in a context of actual brands and a more consequential measure of switching behavior. In an experiment utilizing consumers' realistic experience with a novel mobile app, study 4 then demonstrates the underlying mechanism via resource control's mediating effect and its interaction with contingency on consumers' responses to different violation types, testing our complete conceptual model. Finally, using text analysis of scraped Twitter data, study 5 tests these effects in a naturalistic setting and employs a different measure of control-reclaiming behavior.

Theoretical Framework

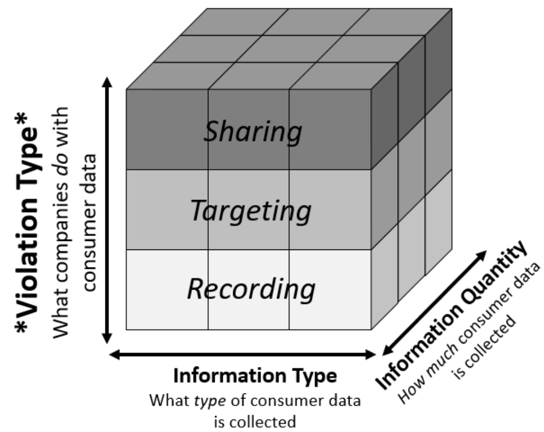
Consumer Privacy and Privacy Violation

Privacy has traditionally been defined as a right "to be left alone" (Brandeis and Warren 1890), specifically with reference to one's own physical space. Due to recent shifts in technology, contemporary definitions of consumer privacy often imply information privacy (Goodwin 1991), especially online (Matz, Appel, and Kosinski 2020). Indeed, the "information age" has created such widespread consumer demand for information privacy that it is sometimes regarded as a commodity that could be regulated through a market structure (Acquisti et al. 2016; Smith et al. 2011). In theory, this suggests that consumers may willingly exchange privacy and personal information for access to products and services (Maseeh et al. 2021).

Despite this theoretical guidance and the growing attention paid to consumer privacy, there remains a lack of consensus as to what constitutes a privacy violation from a consumer perspective (Martin and Murphy 2017). In conceptualizing this construct, the marketing literature has historically focused on the type (Phelps, Nowak, and Ferrell 2000) and quantity of information (Nowak and Phelps 1992) collected by marketers. However, more recent research has recommended a shift away from this focus because associated negative outcomes (e.g., data breech or security lapse) are less frequent and only occur after consumption decisions are made (Martin, Borah, and Palmatier 2017).

In contrast, through the rapid advent of algorithmic targeting and repeated instances of high-profile cases of third-party data sharing, consumers have become keenly aware of and concerned about data usage (Matz et al. 2020). Therefore, consistent with more recent privacy research (Malhotra, Kim, and Agarwal 2004; Martin 2018; Maseeh et al. 2021), we argue that consumers may rely heavily on what the company *does* with their information in determining whether and to what extent a privacy violation has occurred. Further, through a synthesis of research in law, information systems, and consumer psychology, we develop a comprehensive conceptualization of this factor (henceforth "violation type") as a consumer construct. Our conceptual analysis reveals three primary types of privacy violation—*recording*, *targeting*, and *sharing*—that represent a progressive increase in perceived severity. As depicted in figure 2.1, we contend that this focus on violation type compliments existing conceptualizations of privacy violation (Brough and Martin 2020). That is, while the type and quantity of information collected clearly matter to perceived privacy violations, our studies hold these factors constant to permit a clearer focus on the role of violation type as conceptualized here.

FIGURE 2. 1. Conceptualization of Violation Type as Crucial Focus of Privacy Violation Construct



We further theorize that these violation types represent increasing levels of perceived privacy violation (and therefore differentially affect consumer response) because they progressively diminish perceived resource control. That is, because privacy and control are closely linked (Goodwin 1991), we theorize that what companies do with consumers' data influences perceptions of violation severity because it affects the control consumers feel they have over their information. Thus, even when consumers agree to divulge their personal information, the usage of that information may still be perceived as a violation of privacy. Across violation types, *recording* represents the least amount of manipulation over consumer data. Conversely, *targeting* and *sharing* both involve further manipulation and exploitation of such data, suggesting that consumers perceive these violation types as more severe.

Recording. One of the most influential contemporary frameworks for conceptualizing privacy concerns, the Concern for Information Privacy Framework (CFIP; Smith, Milberg, and

Burke 1996), features “collection” as its primary dimension. Smith et al. (1996) define this construct as a “concern that extensive amounts of personally identifiable data are being collected and stored” (p. 172). Together, collecting and storing information constitute *recording*. Research suggests that merely *recording* consumer information—whether through CCTV (Nguyen et al. 2011), computerization of offline behavior (Foxman and Kilcoyne 1993), or online behavioral tracking (Plangger and Montecchi 2020)—is seen by consumers as at least a moderate violation of their privacy. In marketing practice, however, *recording* rarely occurs in isolation but rather serves as a baseline violation that is often accompanied by additional violating actions (Matz et al. 2020). We propose *targeting* and *sharing* as two fundamental violation types that marketers often implement alongside *recording*.

Targeting. Perhaps the most well-known and frequently applied form of privacy violation is *targeting*, represented by a marketer’s use of consumers’ data to deliver personalized advertisements and product promotions (Summers, Smith, and Reczek 2016). In theory this may provide mutual benefit to consumers and marketers by enhancing the relevance of ads that consumers encounter while improving the efficiency with which marketers reach their target segment (Plangger and Montecchi 2020). In practice, however, *targeting* is often perceived by consumers as a pronounced violation of their privacy (Kim et al. 2019) that is more severe than merely *recording* (Matz et al. 2020). This is because, by using consumers’ information in this often unauthorized manner, *targeting* reduces consumers’ perceived control over their information below what it would have otherwise been (Aguirre et al. 2014; Tucker 2014).

Sharing. Although *targeting* and *sharing* are often confounded as “personalization” in the marketing literature (Summers et al. 2016; Vesanen and Raulas 2006), we conceptualize *sharing* as a distinct practice that occurs when the organization that initially recorded consumers’ information subsequently shares that information with another organization. One of the most notable examples of this practice is Facebook's sharing of its users' personal data with Cambridge Analytica, and the scandal that ensued (Confessore 2018). We propose that *sharing* may be seen by consumers as more severe than *targeting*. Gossip theory (Baumeister, Zhang, and Vohs 2004), for instance, suggests that people feel a severe reduction in control over their personal information when it is shared with another party (Martin et al. 2017; Wert and Salovey 2004). More specifically, because *sharing* involves use of consumer data by a third-party, it deprives consumers of control over their information and exposes them to additional risks beyond what occurs through mere *recording* or even *targeting* (Acquisti et al. 2020). Consistent with this theoretical perspective, we expect that consumers will perceive *sharing* (vs. *targeting*) as a more severe privacy violation because it reduces their perceived control over their own information, although both will be perceived as more severe than *recording* for this same reason.

Empirical Validation. The literature, in aggregate, strongly suggests a tripartite conceptualization of privacy violation as outlined above. To supplement our theorizing, we sought empirical validation of the construct through a survey-based, construct-development methodology (Reich and Yuan 2019). The survey was conducted across two samples comprising 252 participants recruited from both an online panel and a university undergraduate participant pool. A fully detailed summary of methods and results is provided in web appendix A.

The survey asked participants to (1) provide their own privacy definition in a text box, (2) provide an example of a privacy violation, and (3) rate their perceived level of privacy violation (1 = *Not at all a privacy violation*; 7 = *An extreme privacy violation*) for each of twelve actual instances of consumer privacy violations. Participants' own provided definition was piped in and repeated for their reference when evaluating each scenario (Reich and Yuan 2019). A PCA conducted on the quantitative data (see table 2.1) and a topic modelling approach (Latent Dirichlet Allocation [LDA]; Blei, Ng, and Jordan 2003) on the qualitative data (see table 2.2) converged to support our privacy violation construct conceptualization, consisting of *recording*, *targeting*, and *sharing* violation types in increasing order of perceived severity. Although the LDA also produced topics related to illegal activity and data breeches, they fall outside the scope of our conceptualization and are therefore disregarded.

TABLE 2. 1. Inductive Study: PCA Results Ordered by Mean Score

	Scenario	Description	Mean	SD	Comp1 3.70*	Comp2 2.04*	Comp3 1.07*
Recording (α=.72)	**Discount**	A retailer offers discounts to members when they scan their mobile barcode from the membership app.	2.54	1.72	0.70		
	Cashier	A cashier asks for customer's zip code at check out	3.19	1.73	0.71		
	Coupon	A supermarket app provides coupons of produce on a customer's phone when s/he around the produce section in-store.	3.78	1.80	0.75		
	CCTV	CCTV in the supermarket aisle that tracks customers' activity in-store.	4.01	1.88	0.65		
Targeting (α=.74)	**History**	A company uses search history to provide ads that are specifically tailored to users.	4.89	1.45		0.76	
	DNAsite	DNA information submitted to an online genealogy website is used to identify a suspect in a criminal case.	4.90	1.98		0.60	
	SMads	A social media app uses users' shared information to provide targeted advertising under sponsored posts.	5.11	1.60		0.77	
	Home-App	A home sharing app uses users' search history to predict high-demand dates and set prices accordingly.	5.13	1.61		0.74	
	PregAds	A retailer uses a customer's purchase history to predict that a customer is pregnant and send pregnancy advertisements to her address.	5.48	1.54		0.70	
Sharing (α=.64)	**Mobile-Service**	A mobile network provider offers a service to other companies to use the provider's customer call and web browsing information to map where customers are located and the types of services they purchase and use	6.04	1.26			0.76
	Access	A social media app allows third-party developers access to users' data.	6.08	1.38			0.61
	Convo	A smart home device accidentally records and sends a conversation of a user to another person.	6.23	1.38			0.84

NOTE.—*component Eigenvalue; numbers in component columns represent factor loadings.

TABLE 2. 2. Inductive Study: Topic Modelling (LDA) Results on Privacy Violation Examples

Topics	Top keywords*	Representative response text**
Recording	monitoring, track, violation, record, permission	“The electronics that we use as a consumer (laptops, phones, cameras). Companies monitoring what you're doing on those machines. Even websites as well” “Google keeping record of what I do on the internet to personalize my page” "surveillance cameras are a violation of privacy because we can't have control over them”
Targeting	Targeting, online, search, ads, history	“The Facebook fiasco. Targeting political ads based on search history.” “The fact that any information I put out in the internet (search and purchase history) is used to generate hundreds of ads on my phone”

		"when you search something online and then see Facebook ads about it. Or when you purchase something and then start getting ads emails."
Sharing	personal, selling, consent, sharing, (third) party(ies)	"Someone sharing my personal information without my consent. Ex: different websites knowing what I am viewing." "Facebook selling your private information to the highest bidder." "Business selling my personal information to third parties without giving me a way to opt out."
Data breach	data, breach, stolen, break, leaked	"A breach happening in which your data is released to the public." "My username and password get leaked in a data breach." "My information was stolen through a data breach once." "A data breach where hackers break into someone's servers and steal my information" "Identify stolen"
Illegal/ physical intrusion	people, hack(-ers,-ing,-ed), card, information, windows	"A customer's information from applying for a credit card is given or misused by the store employee they applied with." "My debit card number was stolen by hackers and someone used it to make purchases." "People peeking into my windows."

NOTE.—topics in bold represent final retained topics from LDA; *words that have higher odds to appear on a certain topic than any other topic; **example responses with probabilities of at least 80% to belong to a certain topic.

Resource Control, Contingency, and Coping

We have thus far conceptualized privacy violation in terms of consumers' perceived control over their personal information, a valued resource. Much like parallel constructs, such as psychological power—the asymmetric control over valued resources in social relations (Fast et al. 2009)—lacking control is generally considered aversive. And the literature suggests that consumers often try to regain control when it is threatened or lost (Whitson and Galinsky 2008). This suggests that, as the severity of privacy violation increases, a decrease in resource control leads to an increased likelihood that consumers will attempt to actively regain control. Indeed, prior research has shown a link between privacy and control (Xu et al. 2012), but extant literature situates the role of control before violations occur. We take a novel perspective in examining control as a consequence of privacy violations and as a core mechanism for subsequent behavior.

Moreover, research has shown that consumers may behave paradoxically with respect to their own privacy. For instance, the literature has previously documented a "privacy paradox" (Aguirre et al. 2014; Krasnova et al. 2010; Sheehan and Hoy 1999) whereby consumers selfreport strong privacy concerns but nonetheless divulge personal information to companies. Our research concerns a different type of privacy paradox observed in the marketplace, in which consumers' privacy is violated yet the associated loss of control sometimes fails to prompt control-reclaiming behavior.

Recent research on psychological control does not provide a satisfactory explanation for this unique, post-violation privacy paradox. Indeed, work by Min and Schwarz (2021) shows that perceptions of low personal influence reduce novelty seeking behavior, whereas perceptions of low predictability of the world enhance novelty seeking behavior. Our conceptualization of privacy violation could theoretically reduce either type of psychological control, and so it remains unclear how existing theorizing might predict switching behavior in this context.

Therefore, to explain this phenomenon we draw from control-adjacent literatures on self-determination and coping. Self-determination theory corroborates our prediction that individuals cope with threats to control by engaging in control-reclaiming behaviors (Deci and Ryan 2000), such as domain-specific attempts to solve a problem (e.g., through brand selection; Schiele and Venkatesh 2016) or as generalized self-affirmations of one's overall sense of control (e.g., through verbal expressions; Thimm, Rademacher, and Kruse 1995). These responses each exemplify a type of active coping. As conceptualized in the clinical psychology literature, active coping occurs when people use their own available resources to reduce the negative impact of a problem (Nicholas, Wilson, and Goyen 1992). This is in contrast to passive coping, which (consistent with Nicholas et al. [1992]) we define as inaction. We expect an active coping response in market situations that provide a high level of contingency (i.e., when consumption outcomes are contingent upon a consumer's own behavior; Weisz 1980; Weisz and Stipek 1982).

Although we focus on brand choice (in studies 1 – 4), study 5 also examines verbal expressions of control affirmation regarding the offending company as an alternate form of active coping (Thimm et al. 1995). Specifically, we draw from research in psycholinguistics on the use of personal pronouns, which have been shown to demonstrate a speaker's mental state (Chung and Pennebaker 2007; Kacewicz et al. 2014). In marketing and related fields, research into personal pronouns has grown in prominence in recent years thanks to the increasing availability of text data (Humphreys and Wang 2018), demonstrating that use of personal pronouns can affect relationships between firms and customers (Packard, Moore, and McFerran 2018) or influence cultural trends (Packard and Berger 2020). With respect to control-reclaiming behavior, Kacewicz et al. (2014) established that use of first-person plural ("we") and second-person singular ("you") pronouns both serve as verbal expressions of control enhancement because they imply a collectivist- and other-orientation, respectively (Cassell et al. 2006). Use of such control-enhancing language can serve as a way for individuals to reclaim control once it has been threatened, even if the individual is not consciously aware of it (Thimm et al. 1995). Therefore, similar to brand switching, we expect greater frequency of control-enhancing pronouns among consumers discussing *sharing* (vs. *targeting*) privacy violations when perceived contingency is relatively high, but not when contingency is low.

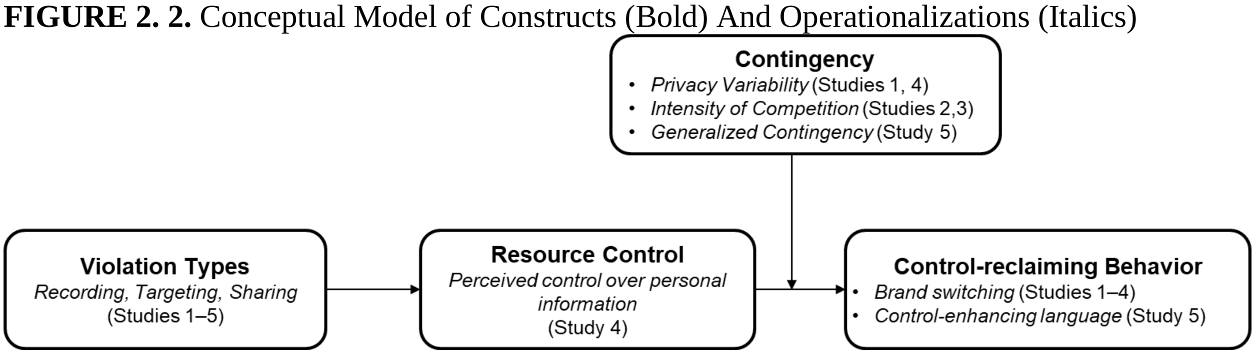
Adding further nuance, the coping literature suggests that consumers may respond passively (through inaction) in market situations in which contingency is low because consumption outcomes are seen as causally disconnected from consumer action. In such cases, a sense of helplessness inhibits control-reclaiming behaviors and no action is taken (Dweck 2000; Skinner 1996), sometimes referred to in a consumer privacy context as a "digital resignation" (Draper and Turow 2019). In practice, we propose that contingency may manifest as industry-level characteristics, such as privacy variability and intensity of competition, that help explain why severe (vs. moderate) privacy violations prompt consumer action in some market contexts (e.g., social media [Hajli and Lin 2016]; search engine [Bleier, Goldfarb, and Tucker 2020]) but not others (e.g., e-commerce; Grosso et al. 2020).

Privacy Variability. We define this characteristic as the amount of variability between companies within an industry in terms of their privacy practices. Much like a “loose monopoly” (Hirschman 1970), an industry with low privacy variability between competitors is likely to be seen by consumers as lacking a meaningful availability of alternatives (Singh 1991). Consumers’ privacy-based choice among brands will likewise seem inconsequential (i.e., low in contingency) and control-reclaiming behavior, even non-choice behavior, will therefore be inhibited. In contrast, high privacy variability should foster a sense of contingency and subsequently control-reclaiming behavior following severe (vs. moderate) privacy violations. This actual marketplace condition, though rarely examined in the literature, has important implications for both marketing practice and theoretical understanding of consumer privacy.

Intensity of Competition. Drawn from the five forces model (Porter 1997), the intensity of competition within an industry may also serve as a manifestation of contingency. Although many industry attributes may contribute to competitiveness (e.g., industry growth rate, exit barriers, etc.), the sheer number of competitors is a primary feature that is most visible to consumers. Interestingly, there need not be an actual dearth of competitors in an industry for consumers to perceive one. Rather, when a significant number of consumers identify with a single supplier, they may perceive a *psychological* monopoly that does not necessarily represent economic reality (Sundie, Gelb, and Bush 2008). For instance, although economic analysis suggests the e-commerce industry is associated with relatively even market-share distribution and high competition (Thomas 2021), many consumers nevertheless perceive Amazon as having a monopoly position in the industry (Jesdanun 2017). Regardless of whether this perception matches economic reality, however, industries perceived by consumers as having a low (high) intensity of competition will likewise create a sense of low (high) contingency, thereby inhibiting (fostering) control-reclaiming behaviors following severe privacy violations.

In sum, we predict that a severe (vs. moderate) privacy violation will prompt greater control-reclaiming behaviors (e.g., brand switching, verbal expressions), but only when the company operates in an industry seen as possessing high contingency characteristics (e.g., high privacy variability, high intensity of competition). This is because, in this market context, regaining resource control seems possible (i.e., contingency is high). However, in a non-contingent industry in which consumers perceive a low privacy variability or intensity of competition, consumers are unlikely to do so regardless of violation severity because regaining control seems impossible (i.e., contingency is low). Importantly, this conceptual model (summarized in figure 2.2) may apply to any context in which consumers surrender their personal information to companies with some degree of uncertainty as to how the data will be used.

FIGURE 2. 2. Conceptual Model of Constructs (Bold) And Operationalizations (Italics)



NOTE.— Measures and attention checks used across studies are included in web appendix B. For brevity and readability, detailed statistics for significant effects are reported in the main text. Further details regarding preregistrations, study designs and ANOVA/regression results are provided in the web appendix.

Study 1

Study 1 examined the moderating role of contingency (operationalized as privacy variability) in consumer control-reclaiming behavior (operationalized as brand switching). The study employed a 3 (violation type: *recording*, *targeting*, *sharing*) × 2 (privacy variability: low, high) full-factorial design. As suggested by our theorizing and preliminary data, we expected a progressive increase in control-reclaiming behavior in response to a *sharing* versus *targeting* versus *recording* privacy violation when the industry is highly contingent (i.e., varied in its privacy practice). However, this effect should be mitigated in an industry lacking privacy variability because it reflects non-contingency.

Method

Study 1 recruited 1,300 US residents from MTurk, five of whom did not complete the study and were dropped from the sample (N = 1,024 after attention check exclusions, M_{Age} = 39.93, SD_{Age} = 12.89; 50.7% female). Participants were given information about "Industry X," containing four companies (A, B, C, and D) and were asked to imagine themselves as current customers of "Company A," the market leader. In the low (high) variability condition, all companies in Industry X used consumer data in the same way (different ways). Participants in the *recording* (*targeting*) [*sharing*] condition then saw a pop-up detailing Company A's cookies policy of recording consumers' data (using consumers' personal data to deliver targeted ads) [sharing consumers' data with third parties]. See appendix A for stimuli.

Brand switching, the core dependent variable, was measured using a dichotomous forced-choice item. Participants indicated whether they would continue onto Company A's website (0) or switch to another brand (1). As a further check of the violation type manipulation, a four-item measure of perceived privacy violation ($\alpha = .84$), adapted from Krasnova et al. (2010), was then administered. To check if the privacy variability manipulation was successful, we then asked participants to rate "How much variety is there in privacy policies of companies in Industry X" on a seven-point scale (1 = *All have the same privacy policies*; 7 = *All have different privacy policies*). Lastly, participants completed an attention check regarding Company A's cookies policy, followed by demographic questions (see web appendix B for measure and attention check details across all studies).

Results

Manipulation Checks. We first conducted a 3 (violation type) × 2 (privacy variability) ANOVA (*df* for *F*-tests = [2, 1018]) on perceived privacy violation. Results showed a significant main effect of violation type on privacy violation ($F = 72.98$, $p < .001$, $\eta_p^2 = .13$; $M_{\text{Recording}} = 3.50$, $SD_{\text{Recording}} = 1.41$, $M_{\text{Targeting}} = 4.00$, $SD_{\text{Targeting}} = 1.31$, $M_{\text{Sharing}} = 4.71$, $SD_{\text{Sharing}} = 1.25$) and no main effect of privacy variability ($p = .92$). We also observed an unexpected violation type × privacy variability interaction ($F = 3.58$, $p = .03$, $\eta_p^2 = .01$). However, the effect of violation type within each level of privacy variability was consistent, displaying both a significant main effect (*p*s < .001, $\eta_p^2{}_{\text{Low Variability}} = .08$, $\eta_p^2{}_{\text{Low Variability}} = .17$ and linear contrast (*p*s < .001) at both levels of privacy variability. Quadratic contrasts were not significant (*p*s > .35). Overall, this confirms that the violation type manipulation increases violation severity in a progressive trend, regardless of privacy variability.

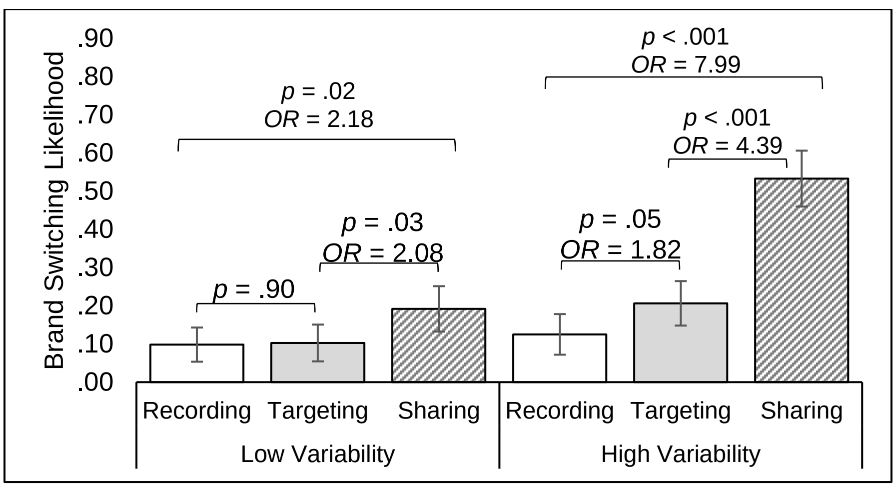
A parallel ANOVA on the variability check revealed the expected main effect of privacy variability ($F = 2997.11$, $p < .001$, $\eta_p^2 = .75$; $M_{\text{High Variability}} = 6.44$, $SD_{\text{High Variability}} = 1.12$, $M_{\text{Low Variability}} = 1.69$, $SD_{\text{Low Variability}} = 1.63$). We also observed an unexpected main effect of violation type ($F = 5.59$, $p = .004$, $\eta_p^2 = .01$; $M_{\text{Recording}} = 3.78$, $SD_{\text{Recording}} = 2.80$, $M_{\text{Targeting}} = 4.21$, $SD_{\text{Targeting}}$

= 2.75, $M_{Sharing}$ = 4.33, $SD_{Sharing}$ = 2.69) and two-way interaction on the variability check (F = 5.34, p =.005, η_p^2 = .01). However, because perceived variability was significantly greater in the high (vs. low) variability condition at all three levels of violation type (ps < .001; see web appendix C for descriptive statistics within each experimental cell), we disregard this interaction and conclude that our manipulations were successful.

*Effects on Brand Switching***.** To test our core predictions, we regressed brand switching on violation type, privacy variability, and their interaction using hierarchical binary logistic regression, where main effects were tested at step 1 and the interaction term was added at step 2. This was done to permit an accurate test of main effects of the manipulations (step 1) absent the influence of the interaction term (included in step 2; Aiken, West, and Reno 1991). Consistent with our theorizing, results showed a significant main effect of violation type on switching behavior (*Wald* χ^2 [2] = 71.58, p < .001). Planned contrasts showed that those in the *sharing* condition (36.7%) were significantly more likely to switch than those in the *targeting* (15.9%, *Wald* χ^2 [1] = 40.95, p < .001, OR = 3.36, $CI_{.95}$ [2.32, 4.87]) and *recording* conditions (11.1%, *Wald* χ^2 [1] = 53.83, p < .001, OR = 4.76, $CI_{.95}$ [3.14, 7.23]). Switching likelihood did not significantly differ between the *recording* and *targeting* conditions (p = .14). There was also a significant main effect of privacy variability on switching behavior such that those in the high (29.6%) versus low variability condition (13.2%) were more likely to switch (*Wald* χ^2 [1] = 24.71, p < .001, OR = 4.39, $CI_{.95}$ [2.78, 6.93]).

Importantly, these main effects were qualified by a significant two-way interaction (*Wald* χ^2 [2] = 9.89, p = .007, figure 2.3). Planned contrasts within the high privacy variability condition revealed that participants were significantly more likely to switch brands in the *sharing* condition (53.3%) as compared to the *targeting* condition (20.6%, *Wald* χ^2 [1] = 40.23, p < .001, *OR* = 4.39, $CI_{.95}$ = 2.78, 6.93), who were in turn more likely to switch than those in the *recording* condition (12.5%, *Wald* χ^2 [1] = 3.88, *p* = .05, *OR* = 1.82, $CI_{.95}$ [1.00, 3.30]). As theorized, in the low privacy variability condition, these effects were significantly attenuated. Participants were still more likely to switch in the *sharing* (19.2%) versus *targeting* (10.3%, *Wald* χ^2 [1] = 4.99, *p* = .03, *OR* = 2.08, $CI_{.95}$ [1.09, 3.95]) and *recording* conditions (9.8%, *Wald* χ^2 [1] = 5.90, *p* = .02, *OR* = 2.18, $CI_{.95}$ = [1.16, 4.08]), although these effects were much smaller than those in the high variability condition as evidenced by odds ratios less than have as large. Further, switching likelihood did not differ between the *recording* and *targeting* conditions (*p* = .90). For additional details, see web appendix C.

FIGURE 2. 3. STUDY 1: Violation Type × Privacy Variability Interaction on Brand Switching



NOTE.—*p*-values represent planned contrast effects; error bars represent 95% CI; *OR* = odds ratio.

Discussion

Findings from study 1 support our theorizing regarding the nuanced effects of companies' usage of consumer data on behavioral responses. Results affirmed that *recording*, *targeting*, and *sharing* consumer data represent distinct privacy violations that are each seen as progressively more egregious and produce a corresponding likelihood in brand switching. This study also demonstrates that contingency may manifest as an industry-level factor (privacy variability) that moderates this relationship, helping to explain consumers' inconsistent privacy behaviors. If the industry has low variability in privacy practice, then different levels of violation severity may not as readily translate into switching behavior as consumers may feel a sense of digital resignation (Draper and Turow 2019). That is, in privacy-standardized industries, contingency is low and a sense of helplessness therefore inhibits consumers' willingness to actively reclaim control regardless of violation severity. In other industries with more varied privacy practice, consumers are more likely to engage in control-reclaiming behavior as a function of privacy violation severity because it still seems possible for them to reclaim control (i.e., contingency is high).

As conceptualized earlier, *recording* serves as a baseline privacy violation by which either *targeting* or *sharing* violations are often accompanied in practice. Having established the three violation types, subsequent studies therefore focus on the distinction between *targeting* and *sharing* to provide a more conservative and streamlined test of the underlying control-based mechanism. This more accurately reflects most market situations, in which *recording* is inherently involved in both *targeting* and *sharing*, but the latter two need not include each other. In other words, *recording* is ubiquitous (Acquisti et al. 2020) but companies differ in terms of whether consumer data is used for *targeting* or *sharing* with third parties.

Study 2

The core aim of study 2 was to replicate the effects on switching behavior observed in study 1 using an alternate operationalization of perceived contingency (i.e., competitiveness) while including an additional control measure (perceived benefit to the company) as a robustness check. The study, pre-registered on AsPredicted.org, therefore employed a 2 (violation type: *targeting*, *sharing*) × 2 (competitiveness: low, high) full-factorial design. Consistent with our theorizing and study 1 results, we expected an increase in control-reclaiming behavior (as brand switching) in response to a *sharing* versus *targeting* privacy violation when the industry is highly competitive because it implies greater perceived contingency. However, the effect should be attenuated in a less competitive industry because it suggests relatively low contingency.

Method

Study 2 recruited 1,100 US residents from MTurk, 20 of whom did not complete the study and were dropped from the sample (N = 978 after attention check exclusions, $M_{\text{Age}} = 39.91$, $SD_{\text{Age}} = 12.28$; 49.3% female). As in study 1, participants were asked to imagine themselves as current customers of "Company A" within "Industry X." Violation type (*targeting* vs. *sharing*) was manipulated as in study 1. In the low (high) competitiveness condition, participants were then informed that Industry X contained two (eight) companies, including Company A.

Brand switching was measured dichotomously as before, and the same four item check of the violation type manipulation ($\alpha = .73$) was included. In addition, a three-item ad-hoc competitiveness scale ($\alpha = .95$) was administered as a check of the competitiveness manipulation: "The level of competition between companies in Industry X is…" (1 = *Very low*/*Not at all intense*/*Not at all competitive*; 7 = *Very high*/*Very intense*/*Extremely competitive*). Next, to be tested as a covariate, participants indicated whether and to what extent Company A's privacy practice benefits consumers or the company: "Company A's cookies policy primarily benefits…" (1 = *consumers*, 7 = *Company A*). This was included as a robustness check because extant research suggests that consumers who view a company's privacy violation as a necessary piece of a market exchange may forego retaliation against the company (Maseeh et al. 2021). Lastly, participants completed an attention check and demographic questions.

Results

Manipulation Checks. A 2 (violation type) × 2 (privacy variability) ANOVA (*df* for *F*-tests = [1, 974]) on perceived privacy violation showed the expected main effect of violation type ($F = 35.51$, $p < .001$, $\eta_p^2 = .04$; $M_{\text{Targeting}} = 4.07$, $SD_{\text{Targeting}} = 1.30$, $M_{\text{Sharing}} = 4.56$, $SD_{\text{Sharing}} = 1.26$), but no effect of competitiveness ($p = .22$) or interaction between factors ($p = .53$). Similarly, a parallel ANOVA on the competitiveness check revealed only the expected main effect of the competitiveness manipulation ($F = 23.35$, $p < .001$, $\eta_p^2 = .02$; $M_{\text{High Competitiveness}} = 5.17$, $SD_{\text{High Competitiveness}} = 1.46$, $M_{\text{Low Competitiveness}} = 4.70$, $SD_{\text{Low Competitiveness}} = 1.61$; other *p*s > .71). Overall, this suggests that both manipulations were independent and successful.

*Effects on Brand Switching***.** Similar to study 1, our core predictions were tested using hierarchical binary logistic regression in which brand switching was regressed on violation type, competitiveness, and their interaction. Results again showed a significant main effect of violation type on switching (*Wald* χ^2 [1] = 17.85, $p < .001$, $OR = 2.21$, $CI_{.95}$ [1.53, 3.19]), such that participants in the *sharing* condition (41.9%) were significantly more likely to switch than those in the *targeting* condition (29.8%). No main effect of competitiveness was observed ($p = .77$).

More importantly, we again observed a significant two-way interaction (*Wald* χ^2 [1] = 4.08, $p = .04$; for detailed results and visualization, see web appendix D). Planned contrasts replicated the pattern observed in study 1. In the high competitiveness condition, participants were significantly more likely to switch brands in the *sharing* condition (47.7%) as compared to the *targeting* condition (29.2%, *Wald* χ^2 [1] = 17.85, $p < .001$, $OR = 2.21$, $CI_{.95} = 1.53, 3.19$). Conversely, in the low competitiveness condition, switching likelihood did not differ between the *sharing* (35.9%) versus *targeting* condition (30.5%; $p = .21$). Moreover, including the benefit check item did not change the direction or significance of any results (see web appendix D), suggesting that our findings are robust to individual differences in perceived benefit tradeoffs when privacy violations occur.

Discussion

Study 2 provided additional support for our conceptual model, further affirming the importance of violation type (i.e., *targeting* vs. *sharing*) to likelihood of brand switching. Importantly, study 2 findings also supported our alternate operationalization of contingency as intensity of competition, showing that its moderating role mirrors that of privacy variability and therefore supporting our control-based model. Industries lacking competition create a perceived psychological monopoly, leading to a low sense of contingency that inhibits control-reclaiming behavior regardless of violation severity. In contrast, highly competitive industries produce a greater sense of contingency and therefore a greater likelihood of engaging in control-reclaiming behavior following a severe (vs. moderate) privacy violation. In the next study, we aim to test our full conceptual model in a more ecologically valid context.

Study 3

Study 3, pre-registered on AsPredicted.org, aimed to further test the role of perceived contingency (as competitiveness) in a context of actual companies and including a behavioral measure of switching. Like study 2, this study employed a 2 (violation type: *targeting*, *sharing*) × 2 (competitiveness: low, high) full-factorial design. Participants were introduced to the study as a survey about search engine usage, and were first asked to report which search engine they use most often (Google, Bing, Yahoo, or DuckDuckGo). They were then presented with information about Google's privacy policy to emphasize its practice of either *targeting* or *sharing* (depending on condition), followed by a manipulation of competitiveness in the search engine industry. A clickable ad for DuckDuckGo, a privacy-protecting search engine platform, was then presented and participants' click-through rate was observed as a behavioral proxy for interest in brand switching. As before, we expected an increase in switching in response to a *sharing* versus *targeting* privacy violation when the industry is described as highly competitive, but not when it is described as less competitive.

Method

Study 3 recruited 1,204 US residents from Prolific (N = 1,072 after attention check exclusions). An additional 61 participants were excluded who indicated DuckDuckGo as their primary search engine, leaving a final sample of 1,011 participants ($M_{\text{Age}} = 34.46$, $SD_{\text{Age}} = 13.66$; 55.1% female). We reasoned that these participants could not meaningfully indicate switching interest toward DuckDuckGo as it was already their brand of choice. Next, the violation type manipulation (*targeting* vs. *sharing*) was presented as a notice of Google's updated privacy policy (see appendix B for stimuli). In the *targeting* (*sharing*) condition, the notice emphasized that Google's primary usage of consumers' personal information was to deliver targeted advertisements (share it with third-party companies). Participants in the low (high) competitiveness condition were then presented with an industry report explaining why the level of competition in the search engine industry is lower (higher) than ever before.

As a behavioral measure of brand switching, participants were then introduced to DuckDuckGo as a privacy-protecting competitor. An image-based ad for DuckDuckGo (see web appendix E) was presented, and participants were offered the opportunity to click on it and try a search in a new tab. The image was programmed into the survey such that their choice of whether to click would be discreetly recorded, thus serving as a behavioral measure of switching akin to click-through rate, a common metric used by digital marketers. Further, because Prolific participants are financially incentivized to finish each study as quickly as possible, the choice of whether to spend time exploring DuckDuckGo was consequential to participants. Following this choice task, the same manipulation checks for violation type ($\alpha = .88$) and competitiveness ($\alpha = .95$) were administered. Lastly, participants completed an attention check and demographics.

Results

Manipulation Checks. A 2 (violation type) × 2 (competitiveness) ANOVA (*df* for *F*-tests = [1, 1,007]) on perceived privacy violation showed the expected main effect of violation type ($F = 25.47$, $p < .001$, $\eta_p^2 = .03$; $M_{\text{Targeting}} = 4.24$, $SD_{\text{Targeting}} = 1.33$, $M_{\text{Sharing}} = 4.65$, $SD_{\text{Sharing}} = 1.29$), but no effect of competitiveness ($p = .39$) or interaction between factors ($p = .40$). Similarly, a parallel ANOVA on the competitiveness check revealed the expected main effect of the competitiveness manipulation ($F = 694.72$, $p < .001$, $\eta_p^2 = .41$; $M_{\text{High Competitiveness}} = 5.84$, $SD_{\text{High Competitiveness}} = 1.35$, $M_{\text{Low Competitiveness}} = 2.88$, $SD_{\text{Low Competitiveness}} = 2.12$), but no effect of violation type ($p = .11$). We did observe an unexpected two-way interaction on the competitiveness check ($F = 3.98$, $p = .05$, $\eta_p^2 = .004$). However, the significant main effect of competitiveness within each level of violation type was consistent at both levels of violation type (*p*s < .001, $\eta_p^2{}_{\text{Targeting}} = .46$, $\eta_p^2{}_{\text{Sharing}} = .35$; see web appendix E for descriptive statistics within each experimental cell). Overall, this confirms that both manipulations were successful.

*Effects on Brand Switching***.** Brand switching was regressed on violation type, competitiveness, and their interaction. Results showed no main effects of either factor (*p*s > .18), but a significant two-way interaction (*Wald* χ^2 [1] = 3.86, $p = .05$; for detailed results and visualization, see web appendix E). Planned contrasts replicated prior results. In the high competitiveness condition, participants were significantly more likely to switch brands in response to *sharing* (36.3%) versus *targeting* (26.7%, *Wald* χ^2 [1] = 5.34, *p* = .02, *OR* = 1.57, $CI_{.95}$ = 1.07, 2.29). Conversely, in the low competitiveness condition, switching likelihood did not differ between the *sharing* (27.3%) versus *targeting* condition (29.3%; *p* = .63).

For robustness, we reran this analysis to include the 61 participants who were dropped for indicating DuckDuckGo as their pre-existing preferred brand. We observed the same pattern of planned contrasts, where violation type influenced switching when competitiveness was high (*Wald* χ^2 [1] = 4.07, *p* = .04, *OR* = 1.47, $CI_{.95}$ = 1.01, 2.14) but not low (*p* = .61). The interaction term, however, became marginally significant (*Wald* χ^2 [1] = 3.16, *p* = .08). This weakened effect is expected and consistent with our model given that those who already prefer the alternative brand could not meaningfully indicate their interest in switching.

Discussion

Study 3 further supported our conceptual model in a context of actual brands and consequential switching behavior. As theorized, contingency (operationalized as intensity of competition) was shown to explain why consumers actively cope in response to severe (vs. moderate) privacy violations in some circumstances but not others. In the next study, we aim to test our full conceptual model in another ecologically valid context.

Study 4

Study 4, pre-registered on AsPredicted.org, aimed to further test our proposed mechanism through the mediating role of resource control within the broader theoretical model depicted in figure 2.2. To enhance ecological validity, we created a user interface for a novel mobile app and asked participants to evaluate it and decide whether or not to adopt the app. This study therefore tests consumers' control-reclaiming behavior through a routine decision to either keep a mobile app or switch to a competing app, while manipulating privacy variability and violation type.

This study also aimed to rule out an alternative explanation. Martin et al. (2017) suggest that consumer vulnerability—perceived susceptibility to harm (Baker, Gentry, and Rittenburg 2005)—explains the impact of data breaches on consumer response. As conceptualized in the literature, consumer vulnerability arises primarily in extreme circumstances that present an imminent threat to safety. In a majority of routine privacy situations, however, consumers are less likely to feel a sense of personal danger, instead sensing a more generalized loss of control over personal information (Draper and Turow 2019). Consistent with recent survey research (Pew 2019) and theoretical distinctions between control and vulnerability (Hill and Sharma 2020), our theorizing therefore suggests that resource control (and not vulnerability) is the primary driving force underlying consumer response to most routine privacy violations.

Method

Sample and Design. Participants were business undergraduate students recruited from a large public university in the US through a participant pool. To more closely match the reality of app markets, we prescreened participants to prohibit those with no interest in the category of the proposed mobile app. This is because consumers who lack interest in a product category are unlikely to be influenced by the type of privacy violation, regardless of contingency (Plangger and Montecchi 2020). This sampling procedure (along with design and measures) was pre-registered. In total, we received 952 responses from potentially interested consumers (N = 866 after attention check exclusions; $M_{Age} = 21.31$, $SD_{Age} = 2.55$; 50.2% female) who were randomly assigned to conditions in a 2 (violation type: *targeting*, *sharing*) × 2 (privacy variability: low, high) between-participants experiment.

Stimuli and Procedure. Participants were first told that they were to evaluate a location service app called CrowdChkr, which informed users how crowded public locations are. Although the app was hypothetical, it was described to participants as an actual app currently under development. As part of the app demonstration process, participants were asked to provide their initials and zip code to simulate the type of personal information consumers would be asked to give to an actual mobile app. Participants were then asked to try a demonstration of the app's user interface, featuring a generic map with a few marked locations (see appendix C). Participants were reassured that this was only a demonstration and not linked to their actual location. When participants clicked on these different locations, details of the current occupancies of the respective places appeared.

To maintain the cover story that the study was mainly concerned with pilot testing a new app, participants were then asked to evaluate their overall impression of the app's main features across five items ($\alpha = .89$; 1 = *useless/bad/negative/unfavorable/boring*, 7 = *useful/good/positive/favorable/interesting*). Following recommendations of Meyvis and Van Osselaer (2018), this composite measure was treated as a covariate in subsequent analyses to enhance statistical power because it (1) correlated strongly with the mediator ($r = .24$, $p < .001$) and dependent variable ($r = –.24$, $p < .001$), and (2) was not influenced by either manipulation (ps > .81). This was especially important in the present study because the constraints of our participant pool did not permit as large a sample as our other online studies.

After completing the demonstration, participants were told that a CrowdChkr account had been automatically created for them with a username piped from the initials and zip code input at the beginning of the study (though in actuality, no such account existed). As ostensible users of the CrowdChckr app, they were then presented with the app's privacy policy. In the *sharing* (*targeting*) condition, the privacy policy notified participants that the app collected information regarding users' contact details and location, and that this information would be shared with a third-party (used to deliver targeted ads). Participants were then provided with additional information as part of the privacy variability manipulation (see appendix C). In the high (low) variability condition, an animated GIF-image informed participants that there are many other location service apps in the industry, and that all of them have the same (a different) privacy policy with respect to users' data.

Two continuous measures were presented, in randomized order, as potential mediators. Resource control was measured using a three-item scale ($\alpha = .87$) adapted from Ajzen (2002). To test consumer vulnerability as a potential alternative explanation, participants completed a three-item ($\alpha = .93$), nine-point measure of vulnerability (Martin et al. 2017). Presentation of these two measures was randomized to appear either before or after the brand switching choice measure. To test discriminant validity, all three multi-item scales were subjected to PCA with Varimax rotation, and results confirm satisfactory discriminant and convergent validity (see web appendix F for PCA results). To operationalize brand switching, participants chose whether to keep their CrowdChkr account and receive an email with a link to download the app (0) or delete their account and switch to a competing app (1). As a check on the privacy variability manipulation, participants were then asked to rate the level of variability in privacy practice in the location service app industry (1 = *All have the same privacy policies*, 9 = *All have the different privacy policies*). Finally, participants responded to an attention check concerning CrowdChkr's privacy policy and provided demographic information.

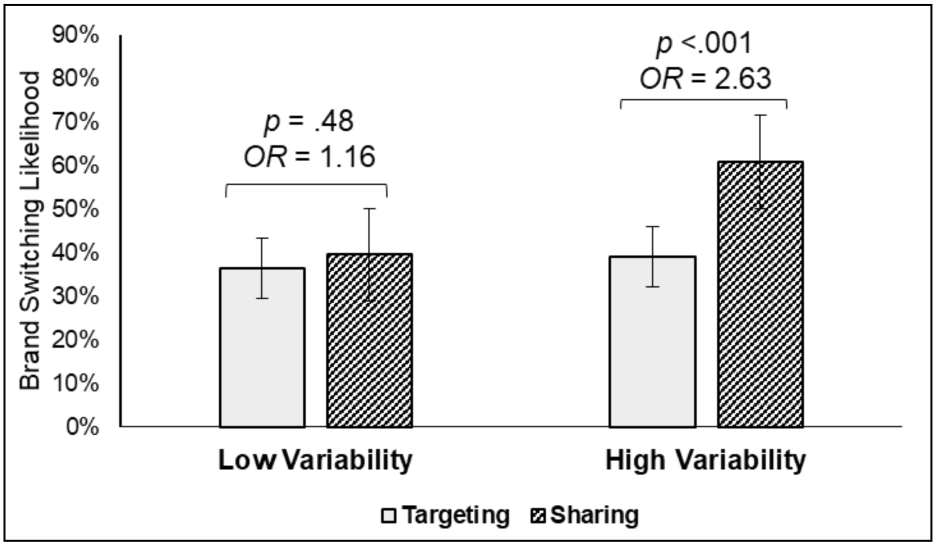
Results

Manipulation Check. A 2 (violation type) × 2 (privacy variability) ANOVA (*df* for *F*-tests = [1, 861]) on the variability check showed no main effect of violation type (p = .22), but an expected main effect of privacy variability such that those in the high (low) variability condition perceived significantly greater privacy variability in the industry (F = 664.33, p < .001, η_p^2 = .44; $M_{\text{High Variability}}$ = 5.21, $SD_{\text{High Variability}}$ = 1.56, $M_{\text{Low Variability}}$ = 2.31, $SD_{\text{Low Variability}}$ = 1.73).

Main Effects. The first part of our conceptual model predicts that a more severe violation type reduces resource control (without influencing vulnerability). As expected, ANOVA results showed a significant main effect of violation type on control (F = 4.74, p = .03; η_p^2 = .01), such that *sharing* (M = 3.26, SD = 1.30) reduced feelings of resource control relative to *targeting* (M = 3.44, SD = 1.31). Privacy variability had no main effect on control (p = .88). A parallel ANOVA on vulnerability revealed no main effects (*p*s > .15), supporting the greater theoretical relevance of control rather than vulnerability.

Interactions. As tested in our earlier studies, our model further predicts that a more severe violation type increases switching likelihood, but only when contingency is high. This was tested by regressing brand switching on violation type, privacy variability, and their interaction using hierarchical binary logistic regression. We observed no main effects (*p*s > .48). However, results supported the expected interaction (*Wald* χ^2 [1] = 7.96, p = .005; figure 2.4). Decomposing this interaction revealed that, when privacy variability was high, participants were significantly more likely to switch brands in the *sharing* (61.0%) versus *targeting* condition (39.1%, *Wald* χ^2 [1] = 22.07, p < .001, *OR* = 2.58, $CI_{.95}$ [1.74, 3.83]). In contrast, when privacy variability was low, there was no significant difference in switching between *targeting* and *sharing* (p = .48).

FIGURE 2. 4. STUDY 4: Violation Type × Privacy Variability Interaction on Brand Switching



NOTE.—*p*-values represent simple contrast effects; error bars represent 95% CI; OR = odds ratio

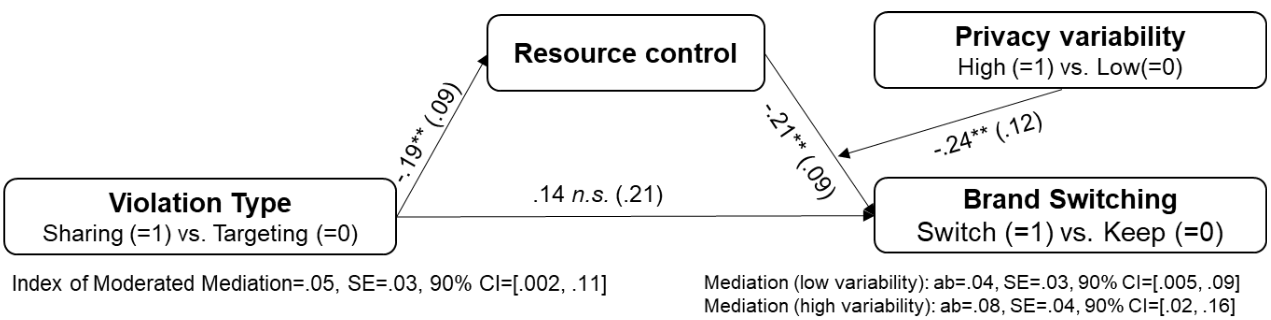
The back end of our conceptual model predicts that variations in resource control translate more strongly into variations in switching behavior when contingency is high (vs. low). Regressing brand switching on resource control, privacy variability, and their interaction provided further support for our theorizing. Control (*Wald* χ^2 [1] = 5.92, p = .02, *OR* = 0.81, $CI_{.95}$ [0.69, 0.96]) and variability (*Wald* χ^2 [1] = 12.31, p < .001, *OR* = 4.31, $CI_{.95}$ [1.91, 9.74]) each had main effects on switching. More importantly, however, we observed the hypothesized interaction (*Wald* χ^2 [1] = 4.90, p = .03; see web appendix F for visualization of interaction and detailed results). Decomposing the interaction revealed that, as predicted, resource control exhibited a strong negative relationship with switching behavior in the high variability condition (*Wald* χ^2 [1] = 34.10, p < .001, *OR* = 0.62, $CI_{.95}$ [0.53, 0.73]) but a significantly attenuated relationship in the low variability condition (*Wald* χ^2 [1] = 4.66, p = .03, *OR* = 0.83, $CI_{.95}$ [0.70, 0.98]) as evidenced by the latter's smaller *Wald* χ^2 statistic and odds ratio much closer to 1. For comparison, we ran a similar logistic regression with vulnerability (instead of control) as a predictor, and found no such interaction with privacy variability (p = .16).

*Moderated Mediation***.** Our full conceptual model (figure 2.2) suggests that severe (vs. moderate) privacy violations increase consumers' active coping behaviors because they reduce resource control, and that this effect is attenuated when contingency (operationalized here as privacy variability) is low. If true, resource control should mediate the relationship between violation type and switching under high privacy variability, but the mediated effect should be attenuated under low variability. This was tested via moderated mediation analysis using the PROCESS Macro (Hayes 2013; Model 15) with violation type as the predictor, resource control the mediator, privacy variability the moderator, and switching the dependent measure (see figure 2.5).

Results were consistent with our theorizing. First, all above-described effects were significant in this model (ps < .05; see web appendix F). Moreover, the conditional indirect effect of violation type on switching through control was significant and pronounced when privacy variability was high (ab = 0.08, SE = 0.04, $CI_{.95}$ [0.008, 0.18]) but attenuated when variability was low ($CI_{.95}$ [0.001, 0.10]). However, the index of moderated mediation fell just outside the threshold of 95% significance (index = 0.05, SE = 0.03, $CI_{.95}$ [–0.002, 0.12]). We therefore reran this model to examine 90% confidence intervals. Indeed, the full model was significant at the 90% level, suggesting a marginally significant moderated mediation effect ($CI_{.90}$ [0.002, 0.11]) such that the indirect effect of *sharing* (vs. *targeting*) on switching through control was pronounced when privacy variability was high ($CI_{.90}$ [0.02, 0.16]) but attenuated when variability was low ($CI_{.90}$ [0.005, 0.09]).

For comparison, we tested an alternative model that substituted vulnerability for resource control as the mediator, and a non-significant index of moderated mediation ($CI_{.90}$ [–0.01, 0.06]) further rules out vulnerability as an alternative explanation. Another similar model was tested treating control and vulnerability as serial mediators (PROCESS model 89). This model showed a marginally significant mediated effect of the interaction on switching through control (ab = 0.04, SE = 0.0, $CI_{.90}$ [0.004, 0.11]), but not through vulnerability (ab = 0.004, SE = 0.01, $CI_{.90}$ [-0.02, 0.03]). Moreover, the overall moderated serial mediation through control and vulnerability in sequence was not significant (index = 0.007, SE = 0.01, $CI_{.90}$ [-0.01, 0.03]). We therefore conclude that control indeed serves as a more robust mechanism than vulnerability. See web appendix F for detailed results of these alternative models.

FIGURE 2. 5. STUDY 4: Moderated Mediation Model



NOTE.—**p* < .10; ***p* < .05;****p* <.01; *NS* = not significant at *p* > .10; coefficients (standard errors) are unstandardized; for clarity, the privacy variability × violation type interaction is omitted from depiction.

Discussion

Study 4 findings further solidify our theorizing. When contingency was high, *sharing* (vs. *targeting*) reduced feelings of control over consumers' own personal information, which in turn increased control-reclaiming behaviors (as brand switching). However, when contingency was low, feelings of control were less likely to translate into control-reclaiming behavior. Furthermore, our results rule out consumer vulnerability as an alternative mechanism, suggesting that consumers' seemingly paradoxical privacy behaviors are primarily rooted in their need to regain lost control. This provides additional support for our theorizing around the roles of resource control and contingency in explaining consumers' seemingly paradoxical privacy behavior. In our final study, we aimed to replicate the violation type × contingency interaction in an actual market setting and using an alternative form of control-reclaiming behavior.

Study 5

Thus far, our studies have shown consistent support for our theorizing in constructed scenarios. To build generalizability, this final study aims to replicate our findings in a more naturalistic setting involving actual brands in distinct industries and employing a directly observable measure of control-reclaiming behavior from a secondary source.

Specifically, study 5 used text analysis of Twitter data to compare linguistic response to *targeting* or *sharing* privacy violations from Amazon and Google (Search), creating a naturalistic 2 (violation type: *targeting, sharing*) × 2 (contingency: high, low) quasi-experiment. Amazon and Google represent two industries (online retail and search engine, respectively) that naturally differ in terms of level of perceived contingency (as shown through a pretest in the next section). We expect increased control-reclaiming behavior in response to a *sharing* (vs. *targeting*) privacy violation from Google because it operates in an industry with high contingency (Ahmad 2018; Norton 2020). Although Google is the market leader, there are significant variations in privacy practices among competing search engines—such as DuckDuckGo (Lomas 2019) and Proton (Edelman 2022)—and economic analysis shows that the industry is highly competitive (Egan 2021). However, because Amazon operates in an industry in which consumers perceive low contingency (i.e., low privacy variability and/or intensity of competition; Grosso et al. 2020; Paul 2020), a sense of helplessness should inhibit consumers' control-reclaiming behavior regardless of the company's privacy-violating action. Consistent with research in psycholinguistics (Chung and Pennebaker 2007), we operationalize control-reclaiming behavior as frequency of "we" and "you" pronouns used in consumers' tweets. We expect greater frequency of these control-enhancing pronouns among consumers tweeting about Google's *sharing* (vs. *targeting*) privacy violations because contingency is relatively high, but no violation-based differences in response to Amazon (where contingency is low).

Pretests

Two pretests were conducted for study 5 (see web appendix G for full details). First, to confirm that Google and Amazon indeed operate in industries with different levels of perceived contingency, 300 MTurk participants (N = 252 after attention check exclusion; $M_{\text{Age}} = 37.83$, $SD_{\text{Age}} = 9.51$; 40.07 % female) were asked to rate perceived level of contingency, using a 7-point ad-hoc scale, regarding both search engine and e-commerce industries. Results of repeated-measure ANOVAs showed that, as theorized, e-commerce was seen as offering significantly less contingency than the search engine industry (*Huynh-Feldt F* (1, 251) = 6.87, $p < .01$, $\eta_p^2 = .03$, $M_{\text{e-commerce}} = 5.27$; $SD_{\text{e-commerce}} = 1.33$, $M_{\text{search}} = 5.45$; $SD_{\text{search}} = 1.12$). Participants also rated market dominance of Amazon and Google in their respective industries, and both were seen as equally dominant (*Huynh-Feldt F* (1, 251) = 1.02, $p = .31$), suggesting that these two companies were well-suited for analysis in the main study.

A second pretest was conducted to validate pronoun use as an operationalization of control-reclaiming behavior. Two hundred Prolific participants (N = 182 after attention check exclusions, $M_{\text{Age}} = 38.64$, $SD_{\text{Age}} = 13.54$, 48.50% female) viewed stimuli about Company A in Industry X, similar to the condition of study 1 representing *sharing* and high privacy variability. Brand switching was measured as in study 1, along with an open-response box prompting participants to write an anonymous tweet (in Twitter format) about Company A. We scraped the open-response text for frequency of “we” and “you” usage, as well as other pronouns. As theorized, brand switching was significantly correlated with we/you frequency (r = .20, p = .01) but not with any other pronoun (ps > .31). Looked at another way, an ANOVA revealed that those who chose to switch (vs. those who chose to stay) used “we” and “you” more frequently (F (1,181) = 4.91, M_{Switch} = 4.80, SD_{Switch} = 5.74, M_{Stay} = 3.03, SD_{Stay} = 4.92, p = .03, η_p^2 = .03), supporting concurrent validity of these verbal expressions as a measure of control-reclaiming behavior to be used in the main study.

Method

We used package Rtweet to collect 6,916 tweets (N = 2,824 after removing duplicates, retweets and replies) from the Twitter Developer API for the one-month period from January 23 to February 22, 2022. These tweets, written in English, were selected because they contained matched keywords (Hewett et al. 2016) reflecting violation type (*targeting* or *sharing*) with either Amazon or Google. Keywords regarding *targeting* (*sharing*) were “targeting,” “targeted,” “targets” (“sharing data,” “shares data,” “shared data,” “third party data,” “3rd party data”). Following Berger et al. (2020), we excluded retweets and replies to ensure that the content was not biased towards the most popular tweets, and removed URLs, punctuation, digits, username tags, and emojis (see table 2.3 for example tweets). We ran the cleaned data through the LIWC Dictionary (Pennebaker et al. 2015) to analyze frequencies (per tweet) of two linguistic expressions of control enhancement: Pronouns “we” (variations include “ours,” “our,” “us”) and “you” (variations include “your,” “yours”).

TABLE 2.3 STUDY 5. Examples of Tweets Collected

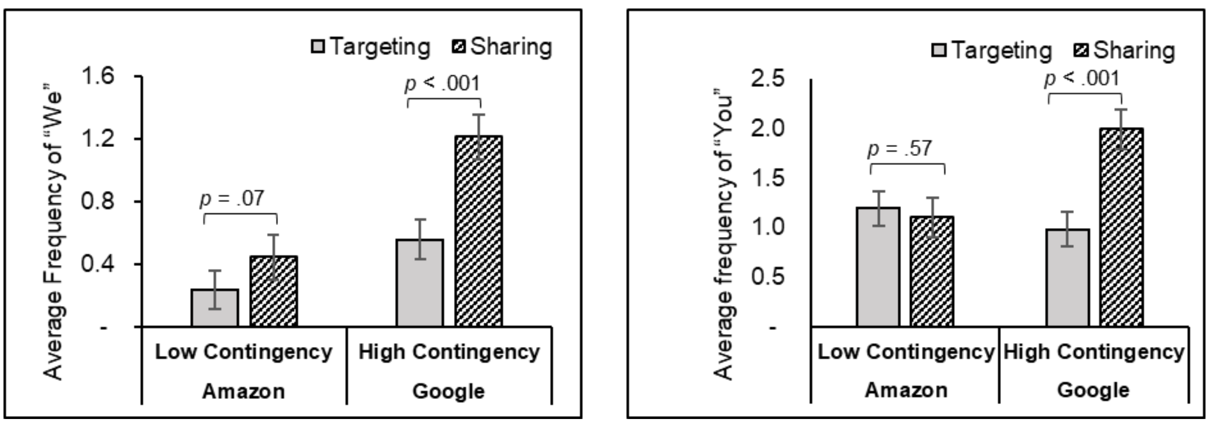
	Amazon	Google
Targeting	"Amazon is LexCorp Reading that the targeted middle and low income audience pays more for Prime than if you can just pay the whole year up front. Targeted. Specifically. Yeah let's milk more from the low income. Apparently, it's a working business strategy" (n= 795)	"the more Google tries to push targeted ads the more enraged i start to feel. i diverge pretty wildly with other more normie libertarian types on this in particular (many who think targeted ads are a consumer good)" (n=929)
Sharing	"Ever wonder why, when you search for a product on flipkart/amazon and the same shit starts appearing in facebook and insta. Its because apps share data with each other." (n=415)	"Google is sharing browsing history and behaviors online, not only sharing it advertisers it is also collecting your data to share with agencies like the NSA and CIA to establish patterns on you and to use against you!" (n=685)

Results

For robustness and clarity, we examine effects on frequency of "we" and "you" both as separate dependent variables and combined as a composite. We remind the reader that both pronouns serve as verbal control-reclaiming behavior and we therefore expect a similar pattern of effects on both outcomes. A 2 (violation type) × 2 (contingency) ANOVA (df = 1, 2820 for F-tests) treating "we" frequency as the dependent variable revealed significant main effects of violation type ($F = 47.78$, $p <.001$, $\eta_p^2 = .02$) and contingency ($F = 44.27$, $p <.001$, $\eta_p^2 = .02$). Results also showed the hypothesized violation type × contingency interaction ($F = 8.69$, $p = .003$, $\eta_p^2 = .003$) and planned contrasts (df = 1, 2820 for F-tests) showed that for tweets referencing Google (figure 2.6, left panel), *sharing* (vs. *targeting*) was associated with significantly greater usage of "we" ($F = 45.68$, $p < .001$, $\eta_p^2 = .02$, $M_{\text{Sharing}} = 1.21$, $SD_{\text{Sharing}} = 2.64$, $M_{\text{Targeting}} = 0.55$, $SD_{\text{Targeting}} = 1.98$), whereas only a marginal effect of violation type was observed among tweets referencing Amazon ($F = 3.19$, $p = .07$, $\eta_p^2 = .001$, $M_{\text{Sharing}} = 0.45$, $SD_{\text{Sharing}} = 1.51$, $M_{\text{Targeting}} = 0.24$, $SD_{\text{Targeting}} = 1.14$).

A parallel ANOVA (df = 1, 2820 for F-tests) on "you" frequency displayed a similar pattern. Planned contrasts again showed, with reference to Google, usage of "you" was significantly greater in response to *sharing* (vs. *targeting*; $F = 56.30$, $p < .001$, $\eta_p^2 = .02$; $M_{\text{Sharing}} = 1.99$, $SD_{\text{Sharing}} = 3.05$, $M_{\text{Targeting}} = 0.98$, $\text{SD}_{\text{Targeting}} = 2.49$), whereas no difference emerged with reference to Amazon ($p = .57$). When we combined *we* and *you* together as a singular dependent variable, the direction and significance of the hypothesized interaction effect and planned contrasts remained (web appendix G). For robustness, we also tested similar ANOVAs treating as outcomes several other pronouns unrelated to control. The hypothesized pattern of effects described above did not emerge with any other pronouns (web appendix G).

FIGURE 2.6. STUDY 5: Violation Type × Contingency Interaction on Frequency of "We" (Left) and "You" (Right)



NOTE.—*p*-values represent planned contrast effects; error bars represent 95% CI.

Discussion

Consistent with our theorizing, study 5 findings suggest that consumers use control-signaling pronouns in a similar manner to brand switching. Specifically, "we" and "you" were used much more frequently when discussing Google's (but not Amazon's) *sharing* (vs. *targeting*) violations. Our theorizing and pretest results suggest that this difference cannot be explained by market dominance, but rather it occurs because Google (vs. Amazon) operates in an industry with greater perceived contingency.

General Discussion

The rapid development of big data analytics (e.g., machine learning, deep learning, artificial intelligence, etc.) suggests that collection and use of consumers' personal data will only continue to grow. Privacy is therefore more prescient to marketers and consumers than ever before (Martin et al. 2017). Yet, consumers continue to patronize the companies that violate their privacy in some cases but not others, presenting a new type of privacy paradox not addressed by extant research. The current research developed and tested a robust model capable of parsimoniously explaining this marketplace phenomenon.

Inspired by several privacy literatures and supported by preliminary data, we first conceptualize three dimensions of privacy violation (*recording*, *targeting*, and *sharing*) that progressively increase in severity (figure 2.1). This tripartite conceptualization serves as a springboard for our broader conceptual model (figure 2.2), based on theories of control, that allows prediction of consumers' seemingly contradictory responses to privacy violations. The model predicts that more severe privacy violations lead to active control-reclaiming behaviors (e.g., brand switching, verbal expressions) because such violations reduce perceived resource control. However, this effect only occurs when market conditions provide a relatively high sense of contingency, such as a high degree of privacy variability or competitiveness in an industry. Four lab experiments and one naturalistic quasi-experiment, conducted across varying samples and contexts, provided support for these predictions.

Theoretical Implications

The current research contributes to theories of consumer privacy by clarifying the construct of privacy violation from a consumer perspective. Because perceived privacy violation is a consumer-level construct, it is crucial to theory development that the construct be examined from the perspective of consumers, rather than academics or practitioners (Reich and Yuan 2019). Our research suggests that consumers may understand privacy violations in a unique way depending on whether the company is *recording*, *targeting* or *sharing* consumer data, even after they have willingly surrendered it (Walker 2016). Importantly, complimenting contemporary conceptualizations (Malhotra et al. 2004; Martin 2018; Sheehan and Hoy 1999), we show that perceptions of privacy violation rely heavily on what companies do with consumers' information, rather than merely the nature or quantity of information collected. In this aspect, we also disentangle the nuanced difference between *targeting* and *sharing*, two violation types traditionally confounded in past literature (Vesanen and Raulas 2006).

In addition, our research establishes the centrality of control and related constructs (e.g., contingency) in shaping consumers' response to privacy violation. Beyond the effects of vulnerability (Martin et al. 2017), our findings suggest that privacy violations reduce perceived resource control, motivating behavioral responses aimed at reclaiming it, such as brand switching. Findings further show that this process is disrupted when market contexts diminish consumers' sense of contingency. Accordingly, we extend theories of consumer privacy by embedding control, at both the industry and consumer level, as a broad theoretical frame for understanding consumer behavior after a violation has occurred. This is especially important in light of the prevalence of consumers' seemingly paradoxical privacy behavior, whereby surface-level antecedents (e.g., privacy concerns) often fail to translate into corresponding behavioral responses (Aguirre et al. 2014; Martin et al. 2017; Norberg, Horne, and Horne 2007). Our research helps explain a novel type of privacy paradox, demonstrating that consumer reactions to privacy violation are primarily rooted in a need to reclaim lost control over personal information, thereby contributing toward a deeper understanding of privacy from a consumer psychology perspective (Krishna 2020). Although control has been examined in this context, it has been examined primarily as antecedent to privacy violations (Spiekermann and Korunovska 2017; Xu et al. 2012) and in a manner that provides little clarity as to why consumers only sometimes act to reclaim control in a privacy context. In contrast, we implement a more comprehensive conceptualization that accounts for the differing post-violation roles of control at multiple levels.

Most importantly, the current research unifies disparate literatures to construct a theoretically-derived and empirically robust model of consumer response to privacy violation. By synthesizing theories of self-determination (Deci and Ryan 2000) and coping (Nicholas et al. 1992) with control, our model shows that people cope passively with threats to control when contingency is low (Skinner and Zimmer-Gembeck 2011) and privacy violations therefore fail to prompt an active consumer response when market conditions make reclaiming control seem impossible. Just as consumers deleted Facebook but continued with Amazon following similar privacy violations, our findings show that consumers are more likely to attempt to reclaim control (through brand switching [studies 1 – 4] or control-enhancing language [study 5]) only in a market context that promotes contingency through privacy variability or competitiveness. In so doing, the current research extends the privacy literature that has previously relied more narrowly on individual differences or economic calculations to explain contradictory consumer response to privacy violations (Bleier et al. 2020; Hallam and Zanella 2017; Norberg et al. 2007).

Relatedly, our research expands the meaning of the "privacy paradox" that has previously been documented in the literature (Acquisti et al. 2020; Norberg et al. 2007). That is, we show that consumers exhibit seemingly paradoxical privacy behaviors even after a violation has occurred. This opens further theoretical avenues for exploring other types of privacy paradoxes, while suggesting that control-adjacent theories may be useful in explaining them.

Practical Implications

As consumer data become more valuable to companies' business models (Interactive Advertising Bureau 2019), marketers must understand how consumers evaluate their data practice and strive to maintain a healthy balance between providing customized experiences and preserving consumer privacy. Our research shows that not all privacy violations are equal and violation severity has meaningful consequences for brands. Specifically, we show that consumers view third-party data *sharing* as a much more severe privacy violation than *recording* and *targeting*. Consequently, companies are more likely to alienate consumers when asking for blanket consent of personal information in vague terms. As our findings suggest, a more balanced approach may be optimal, where companies ask consumers for consent to each data practice separately. For example, a mobile phone company could ask, separately, for consent to (1) record information regarding phone performance and errors, (2) use personal information to deliver targeted offerings, and (3) share certain information with a third party. Such a practice could encourage consumers to continue to engage with the company and share their information as appropriate while preserving consumers' sense of control over their personal information.

In addition, past research has shown that while privacy is important to business practice, it is often seen as a burden or cost (Bamberger and Mulligan 2010; Martin and Murphy 2017). However, our findings suggest that privacy could be used as a unique selling point for a newcomer or a disadvantaged competitor in a saturated market. Indeed, study 3 suggests that emphasizing the competitive nature of the industry (e.g., through advertising) may encourage otherwise reluctant consumers to switch from a dominant brand (e.g., Google) to one that offers consumer-friendly privacy practices, when they sense the brand has violated their privacy. Our other studies provide additional evidence that this can be accomplished in other industries and through other forms of contingency, suggesting privacy-based differentiation opportunities for marketers that may simultaneously enhance profitability and consumer well-being.

Finally, from a policy standpoint, our findings have at least two major implications. First, in establishing the moderating effect of contingency (as privacy variability and competitiveness), our research suggests that consumers are especially susceptible to severe privacy violations in industries with standardized privacy practices or weak competition (e.g., banking or online retail). Consequently, these industries require more attention and regulation from policymakers. Especially when mandating consumer-friendly privacy practices is not feasible, policymakers might instead consider prohibiting privacy policy standardization by mandating a certain level of privacy variability or encourage newcomers and companies with different privacy practices through public funding and legislation, thereby enhancing contingency. Regarding competitiveness, the US congress is currently considering the American Innovation and Online Choice Act, which would prohibit anti-competitive behavior of large tech platforms that often violate consumers' privacy (Romanoff 2022). Our research reinforces the importance of passing such legislation, showing that reduced competition not only has economic consequences, but also inhibits consumers' willingness to act after their privacy has been violated. Second, our findings advocate for public communication and education of consumer privacy options. Our studies show that consumers are susceptible to their *perceived* contingency of a market situation. It is therefore important to remind and reinforce consumers' understanding of their options, encouraging active response towards privacy violations by nurturing perceptions of contingency.

Limitations and Future Research

Although we have attempted to maximize the rigor of our research, several limitations exist that provide opportunities for future research. First, although study 4 adopted a realistic consumption setting and studies 3 and 5 employed naturalistic observation of behavior, none of our studies employed a true field-experiment design in which random assignment of conditions was linked to actual marketplace behavior. One possible approach to doing so would be to deliver manipulated cookies policy pop-ups (similar to studies 1 and 2) during consumer visits to actual company websites, and observing naturalistic responses without their awareness that they are participating in a research study. Although such an endeavor requires ambitious effort, it would be a worthwhile extension of the current research.

In addition, although we have established contingency as an explanatory factor, other factors may contribute to the phenomenon as well. For instance, brand loyalty has been shown to increase consumers' willingness to forgive a company for unethical behavior (Ingram, Skinner, and Taylor 2005) and service failure (Henderson, Beck, and Palmatier 2011). It may therefore also serve as a consumer- and brand-level factor that helps explain why consumers sometimes fail to act on their privacy concerns. Relatedly, because consumers lose certainty when their data is shared, uncertainty may partially explain why *sharing* is seen as the most severe violation type. Likewise, because consumers may perceive *targeting* (but not *sharing*) as providing some benefit, reciprocity may contribute to *targeting* being seen as a less severe privacy violation. To develop a broader model of consumer privacy behavior, future researchers may explore these moderating factors at multiple levels (consumer, company, industry, etc.).

Moreover, our findings are subject to cultural variations in privacy norms. Research has shown that levels of privacy tolerance differ widely across cultures (Markos, Milne, and Peltier 2017). A more collectivist society might be more accepting towards data *sharing* whereas a more individualistic society might react more negatively to the control threat posed by these privacy violations. Future work should investigate these and other issues related to the generalizablity of our findings across cultural contexts.

Finally, across studies, we demonstrate consumers' coping response to a loss of resource control. However, compensatory control theory focuses on *personal* control—one's belief in their ability to achieve desired life outcomes (Landau, Kay, and Whitson 2015)—as a determinant of coping strategies. An outstanding question is, therefore, what is the link between resource control and personal control? Although one might assume that both types of control are always positively associated, situations in which relinquishing control over information (thereby reducing resource control) is necessary to gain access to an app that could help them achieve certain outcomes (thereby enhancing personal control). This interplay between resource control and personal control represents an interesting and important venue for future research, especially in the consumer privacy domain.

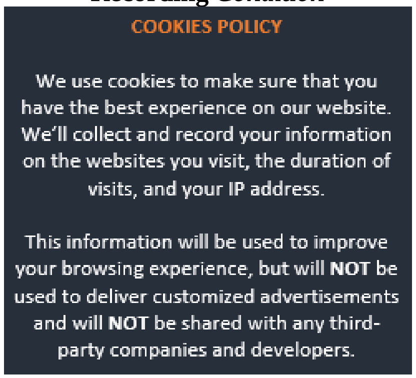
Appendix

Appendix A– Study 1 Stimuli

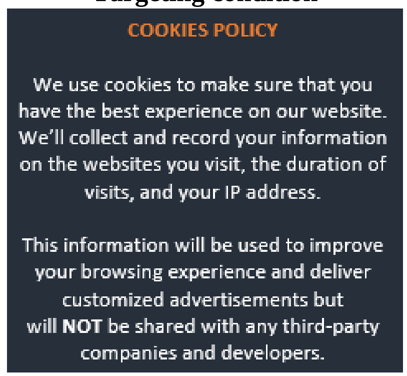
<u>Violation Type Manipulation</u>

You have used this online retailer (social media platform) in the past and are returning today. As you are browsing, and a small pop-up window appears:

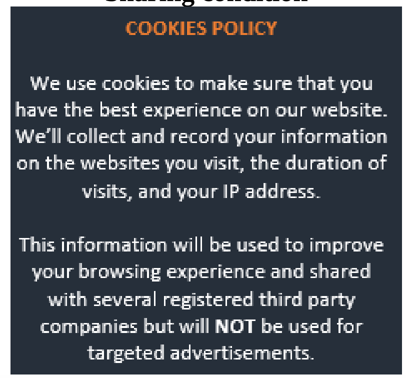
Recording *Condition*



Targeting condition

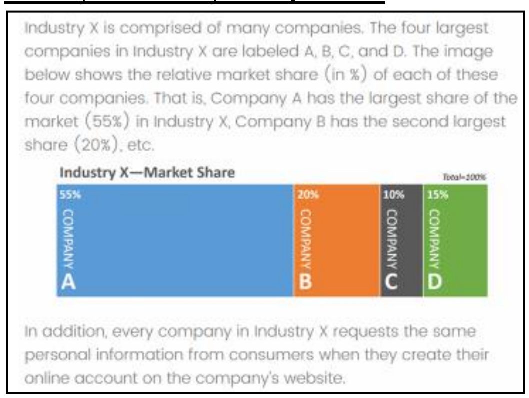


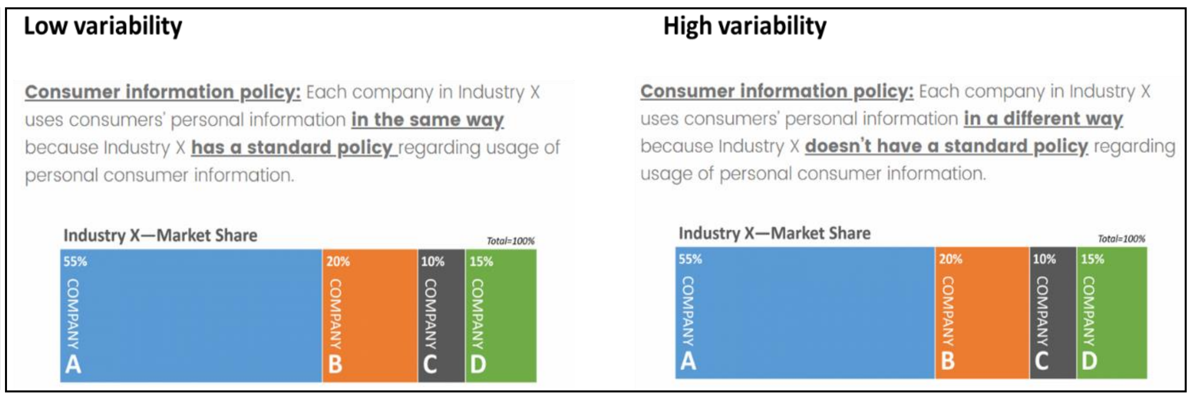
Sharing condition



You recently learned that this cookies practice helps the company earn an annual amount of **$10 per user**.

<u>Privacy Variability Manipulation</u>





Appendix B – Study 3 Stimuli

<u>Violation Type Manipulation</u>

GOOGLE PRIVACY POLICY

When you're interacting with Google Search, we collect some personal information about you.

Targeting Condition
Google's primary use of your personal information (updated): Google uses your personal information to deliver targeted advertisements to you. However, we do not share your personal information with third-party companies and developers, unless you give us permission to do so.

Sharing Condition
Google's primary use of your personal information (updated): Google shares your personal information with third-party companies and developers. However, we do not use your personal information to deliver targeted advertisements to you, unless you give us permission to do so.

By using Google Search, you are accepting the practices described above.

<u>Competitiveness Manipulation</u>

<u>Industry Report: Search Engines</u>

A recent analysis of the search engine industry revealed that the level of competition in the industry is **lower** **(higher)** than ever before, and is continuing to decrease (increase). Google is the dominant search engine, and it essentially holds a monopoly in the industry, preventing other companies from effectively competing for users. This is due to several factors:

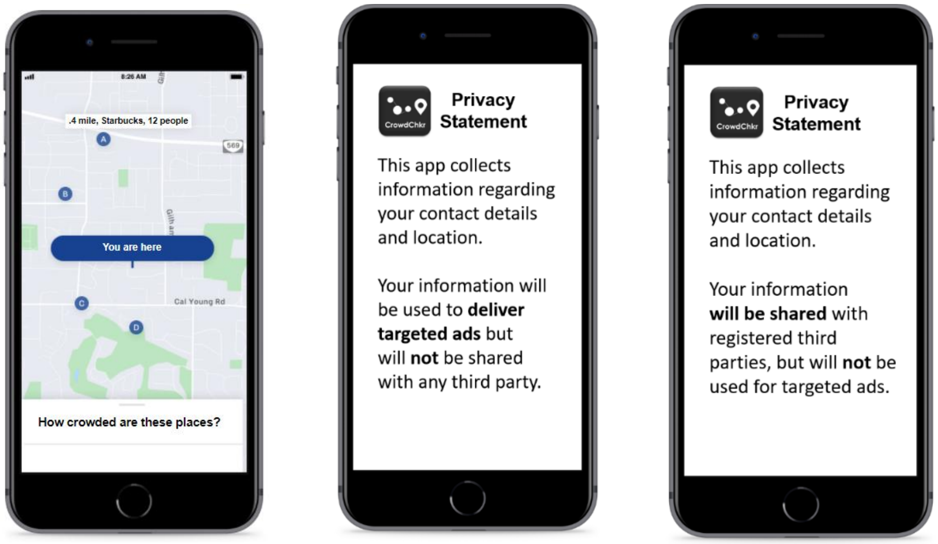
- **Search results** - search engines deliver search results based on proprietary algorithms. Google holds the most advanced algorithm that cannot be imitated by competitors.
- **Barriers to entry** - the search engine industry has high barriers to entry. This means that it is extremely difficult for new companies to enter the industry and compete with Google.
- **Marketing budget** - search engines rely on marketing to attract users. Because Google has a very large marketing budget, it is difficult for other companies with less marketing to compete.

- **Search results** - search engines deliver search results based on proprietary algorithms. Other companies are developing more advanced algorithms that cannot be imitated by Google.
- **Barriers to entry** - the search engine industry has low barriers to entry. This means that it is extremely easy for new companies to enter the industry and compete with Google.
- **Marketing budget** - search engines do not rely on marketing to attract users. Even though Google has a very large marketing budget, it is not difficult for other companies with less marketing to compete.

Note – text highlighted in yellow (blue) was presented in the low (high) competitiveness condition

Appendix C – Study 4 Stimuli

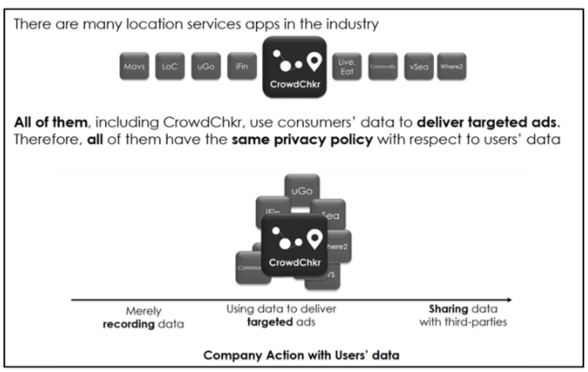
App Interface & Violation Type Manipulation

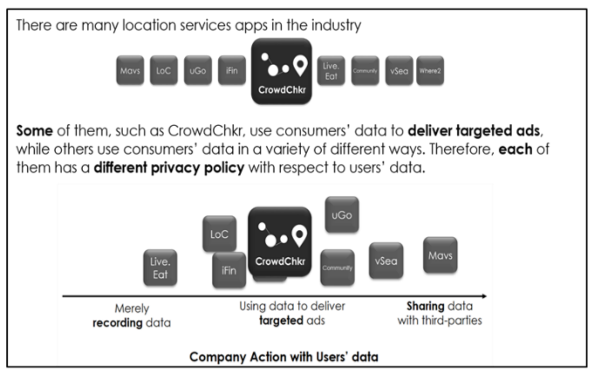


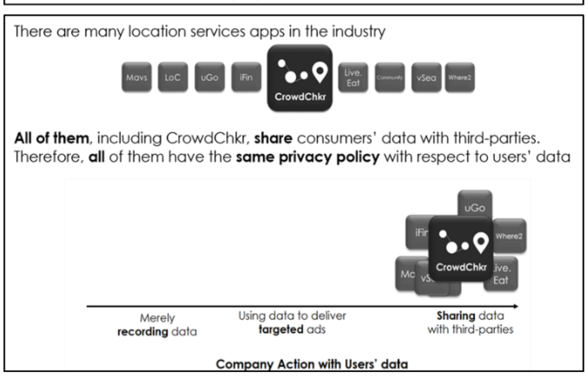
Privacy Variability Manipulation

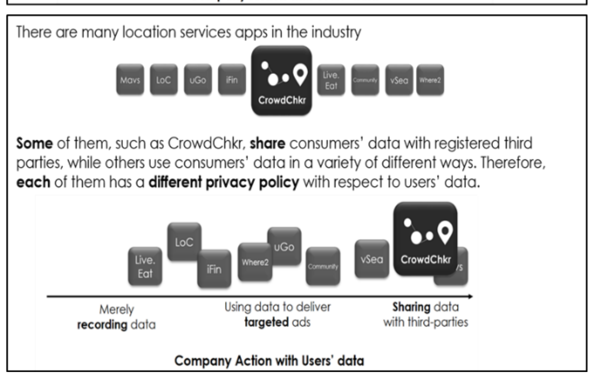
Low Variability

High Variability









Web Appendix

Web Appendix A: Inductive Study

The preliminary validation study utilized several actual instances of consumer privacy violations that have occurred in the past ten years. Participants (N = 252 US residents; M_{Age} = 31.14, SD_{Age} = 1.81; 53.20% female) were recruited from Amazon Mechanical Turk (MTurk; n = 139) and a subject pool of a large public university in the US (n = 113) and were first asked to provide their own privacy definition in a text box. They were then asked to provide an example of a privacy violation. Next, they were presented with twelve scenarios, each presented on its own page and in randomized order (see web appendix A for scenarios). To create these scenarios, we constructed a list of actual privacy incidents reported in the major business press in the past ten years (e.g., The Wall Street Journal, Business Insider, Forbes, The New York Times, etc.) using the business press database Factiva. We excluded incidents of data breaches in which both the company and consumer were victims and filtered out incidents caused by personal errors. We then selected scenarios that garnered substantial media interest at the time and adapted them into short descriptions without mentioning specific companies' names. An incident was said to generate substantial media interest if it was reported by at least five national news and tech news outlets. Each participant's self-reported privacy definition was piped back into a display under each scenario for reference, and participants rated their perceived level of privacy violation for each using a single item: "This action is…" (1 = *Not at all a privacy violation*; 7 = *An extreme privacy violation*).

Principal Component Analysis (PCA).

A PCA with Varimax rotation (table 2W.1) found that the twelve ratings converged into three components with step levels of perceived privacy violation. Component 1 (*recording*; M =

3.34, SD = 1.33) consisted of scenarios in which companies merely collected consumer information (e.g., “CCTV in the supermarket aisle that tracks customers’ activity in-store.”). Component 2 (*targeting*; M = 4.96, SD = 1.21) included situations in which companies used consumer information for targeting purposes (e.g., “A retailer uses a customer’s purchase history to predict that a customer is pregnant and send pregnancy advertisements to her address.”). Component 3 (*sharing*; M = 6.12, SD = 0.99) consisted of actions involving sharing information with third parties (e.g., “A social media app allows third-party developers access to users’ data”).

TABLE 2W. 1. INDUCTIVE—PCA Results Ordered by Mean Score

		Description	Mean	SD	Comp1 3.70*	Comp2 2.04*	Comp3 1.07*
Recording (α=.72)	Discount	A retailer offers discounts to members when they scan their mobile barcode from the membership app.	2.54	1.72	.70		
	Cashier	A cashier asks for customer's zip code at check out	3.19	1.73	.71		
	Coupon	A supermarket app provides coupons of produce on a customer's phone when s/he around the produce section in-store.	3.78	1.80	.75		
	CCTV	CCTV in the supermarket aisle that tracks customers' activity in-store.	4.01	1.88	.65		
Targeting (α=.74)	History	A company uses search history to provide ads that are specifically tailored to users.	4.89	1.45		.76	
	DNAsite	DNA information submitted to an online genealogy website is used to identify a suspect in a criminal case.	4.90	1.98		.60	
	SMads	A social media app uses users' shared information to provide targeted advertising under sponsored posts.	5.11	1.60		.77	
	HomeApp	A home sharing app uses users' search history to predict high-demand dates and set prices accordingly.	5.13	1.61		.74	
	PregAds	A retailer uses a customer's purchase history to predict that a customer is pregnant and send pregnancy advertisements to her address.	5.48	1.54		.70	
Sharing (α=.64)	MobileServ	A mobile network provider offers a service to other companies to use the provider's customer call and web browsing information to map where customers are located and the types of services they purchase and use	6.04	1.26			.76
	Access	A social media app allows third-party developers access to users' data.	6.08	1.38			.61
	Convo	A smart home device accidentally records and sends a conversation of a user to another person.	6.23	1.38			.84

NOTE.—*component Eigenvalue; numbers in component columns represent factor loadings.

A follow-up repeated-measures ANOVA comparing the composite means of these components suggested a significant omnibus difference (Huynh-Feldt $F(1.83, 459.90) = 501.32$, $p < .001$, $\eta_p^2 = .67$), and planned polynomial contrasts showed a significant linear trend ($F(1, 251) = 764.57$, $p < .001$, $\eta_p^2 = .75$). This suggests that consumers perceive progressively increasing levels of privacy violation depending on whether the violation type is *recording*, *targeting*, or *sharing*.

Content Analysis

To further validate our categorization of privacy violation, we conducted follow-up content analyses using both an automated topic modelling technique (Berger et al. 2020; Humphreys and Wang 2018) and conventional thematic coding (Kassarjian 1977) on the examples of privacy violation that participants provided at the beginning of the study. To prepare the data for automated analysis, we first converted all words to lower case and cleaned the data for typos, emojis, URLs, stop words, and punctuations (Berger et al. 2020). We then applied the Latent Dirichlet Allocation (LDA; Blei, Ng, and Jordan 2003) technique to extract potential topics and themes underlying participants' provided example of privacy violations. This approach, similar to PCA, allows groupings of words that frequently occur together into themes or topics, and responses might overlap with the same set of topics in varying probabilities. However, different from the PCA procedure, the LDA algorithm is unconstrained by pre-defined categories and pre-selected scenarios and therefore provides a more naturalistic consumer-generated categorization of privacy violation.

We conducted our analysis with different numbers of topics to generate the set of topics that best balances parsimony with discrimination from one another based on the most frequent keywords. Multiple iterations of the procedure resulted in a final set of five topics, labeled as "recording," "targeting," "sharing," "data breach," and "illegal/ physical intrusion" based on the individual words and responses that represent each topic (table 2W.2). Although the latter two topics are clearly relevant to privacy in general, they represent aberrant violations and are therefore less germane to the current research context of privacy violations routinely levied by marketers in ordinary consumption situations. Consequently, data breach and illegal/physical intrusion were discarded from our conceptualization of privacy violation. The remaining three topics therefore added support for our three-dimension conceptualization of privacy violation faced by consumers.

TABLE 2W. 2. INDUCTIVE STUDY**.** Topic Modelling (LDA) Results on Privacy Violation Examples

Topics	**Top keywords***	**Representative response text****
Recording	monitoring, track, violation, record, permission	"The electronics that we use as a consumer (laptops, phones, cameras). Companies monitoring what you're doing on those machines. Even websites as well" "Google keeping record of what I do on the internet to personalize my page" "surveillance cameras are a violation of privacy because we can't have control over them"
Targeting	Targeting, online, search, ads, history	"The facebook fiasco. Targeting political ads based on search history." "The fact that any information I put out in the internet (search and purchase history) is used to generate hundreds of ads on my phone" "when you search something online and then see facebook ads about it. Or when you purchase something and then start getting ads emails."
Sharing	personal, selling, consent, sharing, (third) party(ies)	"Someone sharing my personal information without my consent. Ex: different websites knowing what I am viewing." "Facebook selling your private information to the highest bidder." "Business selling my personal information to third parties without giving me a way to opt out."
Data breach	data, breach, stolen, break, leaked	"A breach happening in which your data is released to the public." "My username and password get leaked in a data breach." "My information was stolen through a data breach once." "A data breach where hackers break into someone's servers and steal my information" "Identify stolen"
Illegal/ physical intrusion	people, hack(-ers,-ing,-ed), card, information, windows	"A customer's information from applying for a credit card is given or misused by the store employee they applied with." "My debit card number was stolen by hackers and someone used it to make purchases." "People peeking into my windows."

NOTE.—topics in bold represent final retained topics from LDA; *words that have higher odds to appear on a certain topic than any other topic. ** Example responses with probabilities of at least 80% to belong to a certain topic.

The LDA procedure, while useful, may introduce "fuzzy" results due to its unsupervised and automated nature (Humphreys and Wang 2018). We therefore supplemented this analysis by conducting conventional thematic coding. Specifically, two independent coders unfamiliar with the research were recruited to categorize participants' provided examples of privacy violation into these topics using dichotomous coding (1 = related to topic, 0 = unrelated to topic). To simplify the task and focus on the main topics of interest (i.e., *recording*, *targeting*, *sharing*), the coders were instructed to combine the remaining two topics into one group of data breach and illegal/ physical intrusion. To be examined later as a follow-up analysis, we also asked the coders to evaluate participants' definitions of privacy as being related (1) or unrelated (0) to perceived control over personal information. After the coding was completed (inter-rater reliability scores ranging from .81 to .96 across variables), any differences were discussed and resolved to produce a final dataset for comparison with the LDA output.

To test the validity of the LDA procedure, we examined the correlations between the LDA coding results and the corresponding thematic coding results. By nature, for each participant's response, the LDA resulted in a probabilistic mixture of (overlapping) topics while the thematic coding process forced independent topics on the data. Therefore, we did not expect complete matching between the two sets of coding, but rather significant positive correlations only among corresponding topics across the two coding methods. Results from correlation tests supported this prediction (table 2W.3), validating the LDA results and further solidifying our conceptualization of three core privacy violation types.

TABLE 2W. 3. PRELIMINARY STUDY: Correlations between Corresponding Topics of Privacy Violation Examples across Coding Methods

	Thematic Coding Topics			
	Recording	**Targeting**	**Sharing**	**Breach/Illegal/Physical**
Automated Topics (LDA)				
Recording	**.32*** **	.00NS	-.23NS	.07NS
Targeting	-.03NS	**.16 ** **	-.03NS	-.07NS
Sharing	-.07NS	-.04NS	**.52*** **	-.38*
Data Breach	-.13*	-.08NS	-.02NS	.07NS
Illegal/Physical	-.12NS	-.02NS	-.27NS	.33***

NOTE.—coefficients in bold represent the hypothesized correlations; *p <.05, **p < .01; ***p < .001; NS = not significant at p > .05.

Web Appendix B: Measures Used across Studies

Measure	Item	Study
Privacy violation	• "This action is…" (1 = Not at all a privacy violation; 7 = An extreme privacy violation).	Preliminary Study
Privacy violation (Krasnova et al. 2010)	• "This action is…" (1 = *Not at all a privacy violation*; 7 = *An extreme privacy violation*). • "Overall I see no real threat to my privacy due to my participation in this website" (1 = *Strongly disagree*; 7= *Strongly Agree*) • "I feel my personal information is safe on this website" (1 = *Strongly disagree*; 7= *Strongly Agree*) • Overall perception of privacy risk involved when using Company A's website (1 = *Not risky at all*; 7 = *Very risky*)	1, 2, 3
Brand Switching	• In this situation, would you continue on to Company A's website, or would you switch to another company's website to meet your needs? I would… (0 = *Continue with Company A*, 1 = *Switch to another company*)	1, 2, 5-Pretest 2 (5-P2)
Brand Switching	• As previously indicated, using your provided information, a CrowdChkr account was automatically created for you with the username [username] How would you like to proceed with your CrowdChkr account? Please select one of the three options below? Regardless of your choice, the length of this survey and your credit for this study will NOT be affected. • I would… (0 = *Keep my account on CrowdChkr*, 1 = *Delete my account on CrowdChkr and switch to another similar app*)	4
Click Through	Part of this research is also aimed at providing consumers information about other companies within the search engine industry. Below is an example, DuckDuckGo. DuckDuckGo is a search engine that provides users with greater privacy protection while searching. **If you are interested in trying a search on DuckDuckGo** to see how it works, you may click anywhere on the image below. This will open a new tab where you can try it and see for yourself. You will still need to return to this tab to complete the survey. **If you are NOT interested in trying a search on DuckDuckGo**, you may continue on to the next page of this survey **without** clicking the image below. Your decision of whether or not to try a search on DuckDuckGo will not affect your payment for this survey. • Behavioral DV (Click though =1 or not =0)	3

Resource Control (Ajzen 2002)	• How much control do you feel over your information on CrowdChkr? (1 = *No control at all over my information*; 7 = *Complete control over my information*)
	• How much do you agree with the following statements?
	  ○ I feel in complete control over my personal information on company A's website (1 = *Strongly disagree*, 7 = *Strongly agree*)
	  ○ It would be extremely easy for me to control my personal information on Company A's website (1 = *Strongly disagree,* 7 = *Strongly agree*)
		4

Vulnerability (Martin, Borah, and Palmatier 2017)	• How do the personal data practices in this situation make you feel:
	  ○ vulnerable (1 = *Not at all vulnerable*; 7 = E*xtremely vulnerable*)
	  ○ exposed (1 = *Not at all exposed*; 7 = *Extremely exposed*)
	  ○ threatened (1 = *Not at all threatened*; 7 = *Extremely threatened*)
		4

Variability Check	• How much variety is there in privacy policies of companies in Industry X? (1 = *All have the same privacy policies*; 7 = *All have different privacy policies*)	1, 4

(continued in the next page)

(continued from the previous page)

Measure	Item	Study
App Rating	• Please rate the CrowdChkr app using the scales below: ○ 1 = *Useless*; 7 = *Useful* ○ 1 = *Bad*; 7 = *Good* ○ 1 = *Negative*; 7 = *Positive* ○ 1 = *Unfavorable*; 7 = *Favorable* ○ 1 = *Boring*; 7 = *Interesting*	4
Perceived contingency	• Please indicate the extent to which you agree or disagree with the following statements about the e-commerce/search industry ○ Consumers' privacy in the e-commerce/search industry depends upon their choice of brand (1 = *Strongly disagree*, 7 = *Strongly agree*) • Consumers in the e-commerce/ search industry are able to protect their own privacy through the brands they choose (1 = *Strongly disagree*, 7 = *Strongly agree*)	5-P1
Perceived dominance	• In the search engine/e-commerce industry, Google/Amazon is... ○ *clearly not the market leader* (=1) : *clearly the market leader* (=7) *not at all dominant* (=1) : *extremely dominant* (=7)	5-P1
Perceived benefit check	• Company A's cookies policy primarily benefits…(1=Consumers, 7= Company A) • Who benefits most from Google's privacy policy (described above), users or the company? (users = 1; company = 7)	1,2
Tweet	• Open-ended: You are a current customer of Company A. Given the information about Company A and Industry X (repeated below for your reference), what would you write on social media about Company A if you had the chance? Please indicate your response at the bottom of this page	5-P2
Competitiveness check	• The level of competition between companies in Industry X/ Search Engine is… ○ 1 = *Very low*, 7 = *Very High* ○ 1 = *Not at all intense*, 7 = *Very Intense* ○ 1 = *Not at all competitive*, 7 = *Extremely competitive*	2,3
Attention check	• Please select the circle exactly in the middle of the scale for this question. • Please select Strongly Disagree for this question • Company A collects and uses personal data to …deliver targeted ads/ share with registered third parties/ neither of the above • According to the information provided in this study, Google's primary use of users' personal information is to... deliver targeted ads/ share with registered third parties/ neither of the above • Please select Disagree for this question • Which of the following industries was NOT mentioned in this study? (1= *Eye care*; 0= otherwise) • In Industry X, company A has…. the largest/ smallest market share/ don't remember • Which of the following is true about company A? (1= *Company A shares consumers' personal data with third parties*, 0= otherwise)	Pre Study 1,2 2 3 4 5-P1 5-P2 5-P2

Web Appendix C: Detailed Results of Study 1

Study Design

3 (violation type: *recording*, *targeting*, *sharing*) × 2 (privacy variability: low, high) full-factorial design.

Sample

N = 1,295 US residents from MTurk (We planned a sample of 1300, but 5 responses were not finished and not recorded in the data set, N = 1,024 after attention check exclusions, M_{Age} = 39.93, SD_{Age} = 12.89; 5.7% female)

TABLE 2W. 4. STUDY 1: Number of Participants across Conditions (Cell Sizes)

	Variability		
Violation Type	Low	High	**Total**
• **Recording**	173	152	**325**
• **Targeting**	156	189	**345**
• **Sharing**	172	182	**354**
Total	**501**	**523**	**1024**

Manipulation Checks

TABLE 2W. 5. STUDY 1: 3 (Violation Type: Recording, Targeting, Sharing) × 2 (Privacy Variability: High, Low) Factorial ANOVAS on Perceived Privacy Violations & Variability Check

	DV = Perceived Privacy Violation				**DV = Variability Check**			
Variables	*df*	*F*	*p-value*	η_p^2	*df*	*F*	*p-value*	η_p^2
Variability	1	.01	.92	.00	1.00	2,997.11	<.001	.75
ViolationType	2	72.98	<.001	.13	2	5.59	.004	.01
Variability × ViolationType	2	3.58	.03	.01	2	5.34	.005	.01
Error	1,018				1,018			
Total	1,024				1,024			

TABLE 2W. 6. STUDY 1: Estimated Marginal Mean of Perceived Privacy Violation and Variability Check Across Conditions

Condition	**DV = Perceived Privacy Violation**				**DV = Variability Check**			
	Mean	SE	95% CI LL	95% CI UL	Mean	SE	95% CI LL	95% CI UL
Variability = Low								
• **Recording**	3.63	.10	3.43	3.82	1.44	.11	1.23	1.65
• **Targeting**	4.02	.11	3.81	4.23	1.54	.11	1.33	1.76
• **Sharing**	4.57	.10	4.38	4.77	2.09	.11	1.88	2.29
Variability = High								
• **Recording**	3.36	.11	3.15	3.57	6.45	.11	6.23	6.67
• **Targeting**	3.99	.10	3.80	4.18	6.41	.10	6.22	6.61
• **Sharing**	4.84	.10	4.65	5.04	6.45	.10	6.24	6.65

Main Results

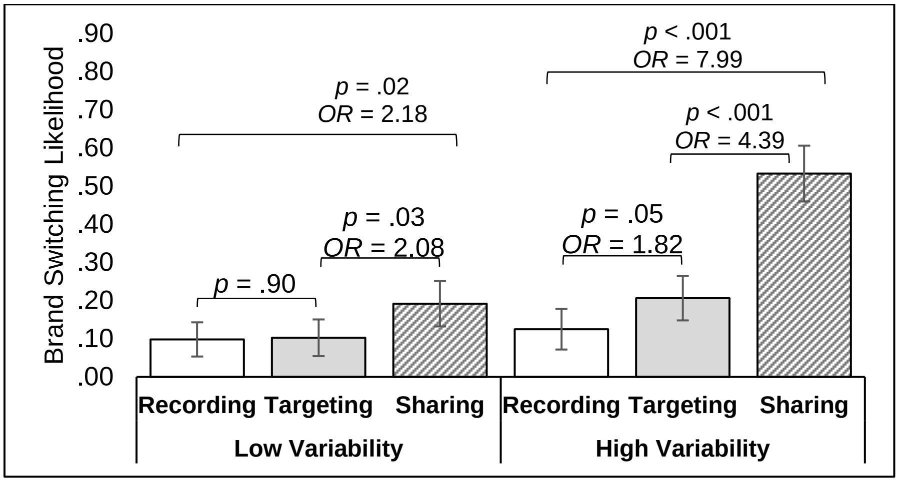
TABLE 2W. 7. STUDY 1: Wald Tests of Hierarchical Linear Logistic Regression of Violation Type and Privacy Variability on Brand Switching

Variables	*Wald* χ^2	df	p-value
Step 1: Main effect of Violation Type and Variability			
• **Variability**	4.51	1	<.001
• **Violation Type**	71.58	2	<.001
Step 2: Interaction effects of Violation Type & Variability			
• **Variability**	41.38	1	<.001
• **ViolationType**	8.04	2	.018
• **Variability × ViolationType**	9.89	2	.007

TABLE 2W. 8. STUDY 1: Tests of Simple Effects of Violation Types at Different Levels of Privacy Variability on Brand Switching

Planned Contrasts	B	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Overall Main effect of Violation Types (collapsed across both conditions of variability								
• **Targeting v. recording**	.35	.23	2.24	1	.135	1.42	.90	2.24
• **Sharing v. recording**	1.56	.21	53.53	1	<.001	4.77	3.14	7.23
• **Sharing v. targeting**	1.21	.19	4.95	1	<.001	3.36	4.87	2.32
Variability = Low								
• **Targeting v. recording**	.05	.37	.02	1	.897	1.05	.51	2.15
• **Sharing v. recording**	.78	.32	5.90	1	.015	2.18	1.16	4.08
• **Sharing v. targeting**	.73	.33	4.99	1	.026	2.08	3.95	1.09
Variability = High								
• **Targeting v. recording**	.60	.30	3.88	1	.049	1.82	1.00	3.30
• **Sharing v. recording**	2.08	.29	52.51	1	<.001	7.99	4.55	14.01
• **Sharing v. targeting**	1.48	.23	4.23	1	<.001	4.39	6.93	2.78

FIGURE 2W. 1. STUDY 1: Violation Type × Privacy Variability Interaction on Brand Switching



NOTE.—*p*-values represent planned contrast effects; *OR*: Odd Ratios; error bars represent 95% CI.

Web Appendix D: Detailed Results of Study 2

Study Design & Preregistration

This study employed a 2 (violation type: *targeting*, *sharing*) × 2 (competitiveness: low, high) full-factorial design

This study was pre-registered on AsPredicted.org

Sample

Study 2 recruited 1,100 US residents from MTurk, 20 of whom did not complete the study and were dropped from the sample (N = 978 after attention check exclusions, $M_{Age} = 39.91$, $SD_{Age} = 12.28$; 49.3% female)

TABLE 2W. 9. STUDY 2: Number of Participants across Conditions (Cell Sizes)

Violation type	**Competitiveness**		**Total**
	Low	**High**	
• **Targeting**	243	260	503
• **Sharing**	234	241	475
Total	**477**	**501**	**978**

Manipulation checks

TABLE 2W.10. STUDY 2: 2 (Violation Type: Targeting, Sharing) × 2 (Competitiveness: High, Low) Factorial ANOVAs on Perceived Privacy Violations & Variability Check

	DV = Perceived Privacy Violation				DV = Competitiveness Check			
Variables	*df*	*F*	*p-value*	η_p^2	*df*	*F*	*p-value*	η_p^2
Competitiveness	1	1.49	.222	.002	1	23.35	<.001	.023
ViolationType	1	35.51	<.001	.035	1	.14	.709	<.001
Variability × ViolationType	1	.40	.526	<.001	1	.01	.943	<.001
Error	974				974			
Total	978				978			

TABLE 2W. 11. STUDY 2: Estimated Marginal Mean of Perceived Privacy Violation and Competitveness across Conditions

Condition	DV = Perceived Privacy Violation				DV = Variability Check			
	Mean	SE	95% CI LL	95% CI UL	Mean	SE	95% CI LL	95% CI UL
Competitiveness = Low								
• **Targeting**	4.04	.08	3.88	4.20	4.72	.10	4.52	4.91
• **Sharing**	4.48	.08	4.32	4.64	4.67	.10	4.48	4.87
Competitiveness = High								
• **Targeting**	4.09	.08	3.94	4.25	5.19	.10	5.00	5.37
• **Sharing**	4.63	.08	4.47	4.79	5.16	.10	4.96	5.35

Main Results

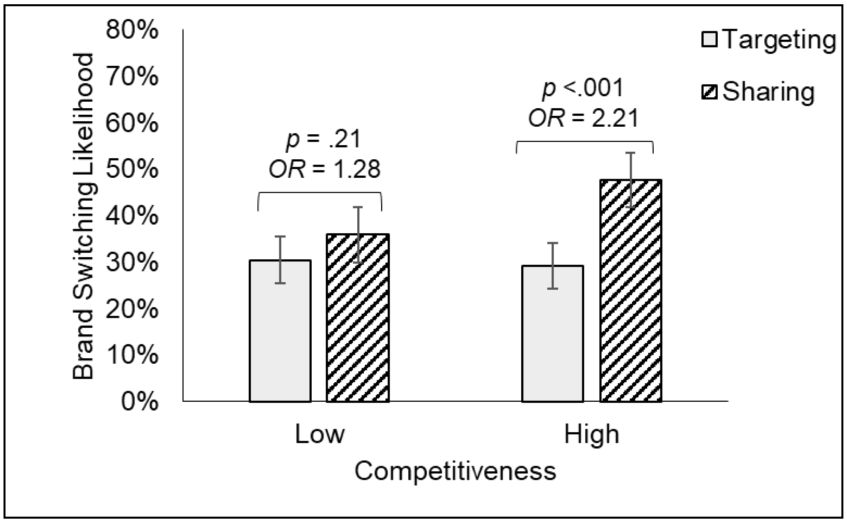
TABLE 2W. 12. STUDY 2**:** Wald Tests of Hierarchical Linear Logistic Regression of Violation Type and Competitiveness on Brand Switching

Variables	*Wald* χ^2	df	p-value
Step 1: Main effect of Violation Type and Competitiveness			
• **Competitiveness**	2.83	1	.093
• **Violation Type**	15.57	1	<.001
Step 2: Interaction effects of Violation Type & Competitiveness			
• **Competitiveness**	.09	1	.765
• **ViolationType**	1.59	1	.207
• **Competitveness × ViolationType**	4.08	1	.043

TABLE 2W. 13. STUDY 2: Tests of Simple Effects of Violation Types at Different Levels of Competitiveness on Brand Switching

Planned Contrasts	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Overall Main effect of Violation Types (collapsed across both conditions of competitiveness								
• **Sharing v. targeting**	.53	.13	15.57	1	<.001	1.70	1.31	2.22
Competitiveness = Low								
• **Sharing v. targeting**	.25	.19	1.59	1	.207	1.28	.87	1.87
Competitiveness = High								
• **Sharing v. targeting**	.79	.19	17.85	1	<.001	2.21	1.53	3.19

FIGURE 2W. 2. STUDY 2: Violation Type × Competitveness Interaction on Brand Switching



NOTE.—*p*-values represent planned contrast effects; *OR*: Odd Ratios; error bars represent 95% CI.

TABLE 2W. 14. STUDY 2: Robustness Check — Logistic Regression of Violation Type and Competitiveness on Brandswitching with Perceived Benefit as Covariate

Variables	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Competitiveness	-.12	.20	.37	1	.543	.88	.60	1.31
Violation Type	.07	.20	.11	1	.735	1.07	.72	1.59
Competitiveness × ViolationType	.58	.28	4.35	1	.037	1.79	1.04	3.09
Perceived Benefit	.53	.07	5.51	1	<.001	1.69	1.47	1.96

Web Appendix E: Detailed Results of Study 3

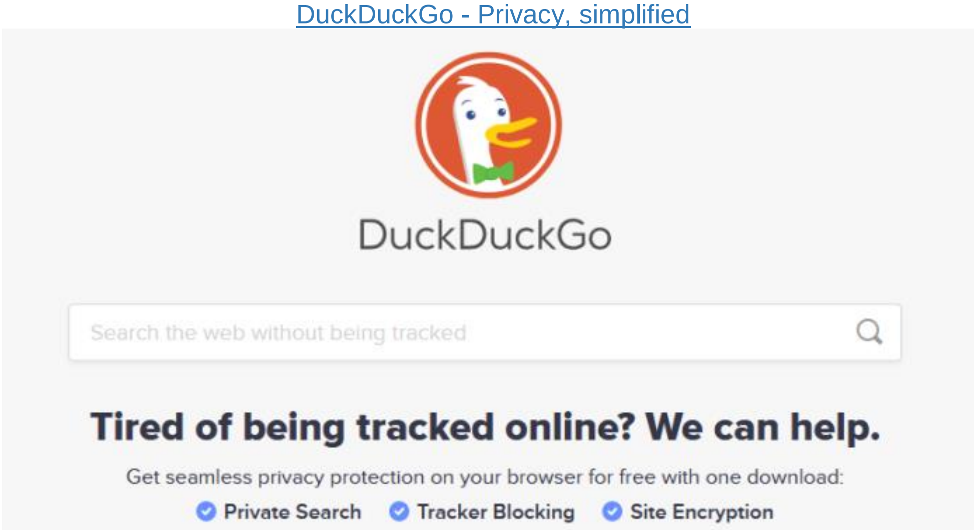
Study Design & Preregistration

This study employed a 2 (violation type: *targeting*, *sharing*) × 2 (competitiveness: low, high) full-factorial design.

This study was pre-registered on AsPredicted.org.

DuckDuckGo's Ad

Part of this research is also aimed at providing consumers information about other companies within the search engine industry. Below is an example, DuckDuckGo.
DuckDuckGo is a search engine that provides users with greater privacy protection while searching.
If you are interested in trying a search on DuckDuckGo to see how it works, you may click anywhere on the image below. This will open a new tab where you can try it and see for yourself. You will still need to return to this tab to complete the survey.
If you are NOT interested in trying a search on DuckDuckGo, you may continue on to the next page of this survey **without** clicking the image below.
Your decision of whether or not to try a search on DuckDuckGo will not affect your payment for this survey.
DuckDuckGo - Privacy, simplified



Sample

Study 3 recruited 1,204 US residents from Prolific (N = 1072 after attention check exclusions, $M_{Age} = 34.50$, $SD_{Age} = 13.74$; 54.15% female)

TABLE 2W. 15. STUDY 3: number of participants across conditions (cell sizes)

Violation type	Competitiveness Low	Competitiveness High	Total
• **Targeting**	291	257	548
• **Sharing**	248	276	524
Total	**539**	**533**	**1072**

Manipulation checks

TABLE 2W. 16. STUDY 3: 2 (Violation Type: Targeting, Sharing) × 2 (Competitiveness: High, Low) Factorial ANOVAs on Perceived Privacy Violations & Variability Check

	DV = Perceived Privacy Violation				DV = Competitiveness Check			
Variables	*df*	*F*	*p-value*	η_p^2	*df*	*F*	*p-value*	η_p^2
Competitiveness	1	1.00	.318	.001	1	74.27	.000	.409
ViolationType	1	22.52	.000	.021	1	3.52	.061	.003
Variability × ViolationType	1	2.17	.141	.002	1	1	4.06	.044
Error	1068				1068			
Total	1072				1072			

TABLE 2W. 17. STUDY 3: Estimated Marginal Mean of Perceived Privacy Violation and Competitveness across Conditions

Condition	DV = Perceived Privacy Violation				DV = Variability Check			
	Mean	SE	95% CI LL	95% CI UL	Mean	SE	95% CI LL	95% CI UL
Competitiveness = Low								
• **Targeting**	4.35	.08	4.19	4.50	2.68	.10	2.47	2.88
• **Sharing**	4.61	.09	4.45	4.78	3.10	.11	2.88	3.32
Competitiveness = High								
• **Targeting**	4.31	.08	4.14	4.47	5.85	.11	5.63	6.07
• **Sharing**	4.82	.08	4.66	4.97	5.84	.11	5.63	6.04

Main Results

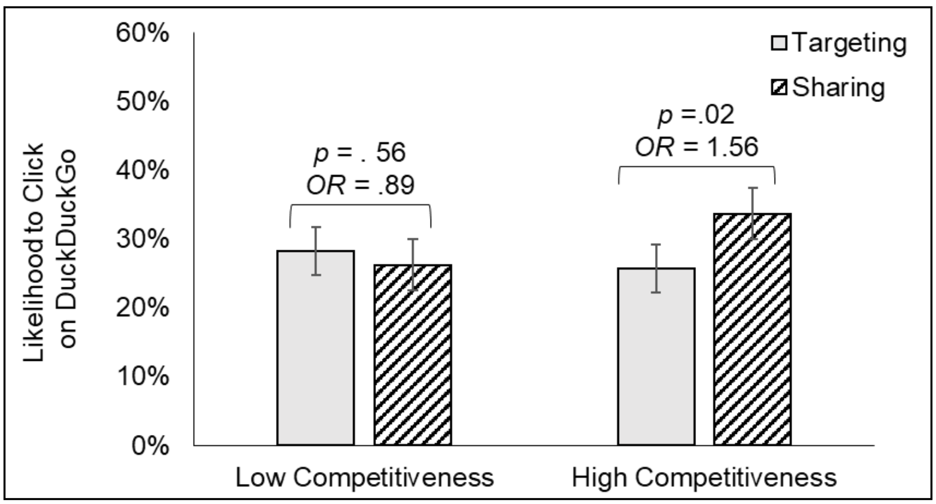
TABLE 2W. 18. STUDY 3: Wald Tests of Hierarchical Linear Logistic Regression of Violation Type and Competitiveness on Clickthrough

Variables	***Wald*** χ^2	**df**	**p-value**
Step 1: Main effect of Violation Type and Competitiveness			
• **Competitiveness**	1.02	1	.313
• **Violation Type**	1.50	1	.220
Step 2: Interaction effects of Violation Type & Competitiveness			
• **Competitiveness**	.52	1	.469
• **ViolationType**	.34	1	.558
• **Competitveness × ViolationType**	4.08	1	.043

TABLE 2W.19. STUDY 3: Tests of Simple Effects of Violation Types at Different Levels of Competitiveness on Clickthrough

Planned Contrasts	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Overall Main effect of Violation Types (collapsed across both conditions of competitiveness								
• **Sharing v. targeting**	.17	.14	1.50	1	.220	1.18	.90	1.55
Competitiveness = Low								
• **Sharing v. targeting**	-.11	.20	.34	1	.558	.89	.61	1.31
Competitiveness = High								
• **Sharing v. targeting**	.44	.19	5.21	1	.022	1.56	1.06	2.27

FIGURE 2W.3. STUDY 3: Violation Type × Competitveness Interaction on Clickthrough



NOTE.—*p*-values represent planned contrast effects; *OR*: Odd Ratios; error bars represent 95% CI.

TABLE 2W.20. STUDY 3: Robustness Check — Logistic Regression of Violation Type and Competitiveness on Clickthrough with Perceived Benefit as Covariate

Variables	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Competitiveness	.15	.20	.57	1	.449	1.16	.79	1.70
Violation Type	.42	.19	4.74	1	.029	1.53	1.04	2.24
Competitiveness × ViolationType	-.57	.28	4.19	1	.041	.57	.33	.98
Perceived Benefit	.06	.06	1.09	1	.297	1.06	.95	1.19
DuckDuckGo_User	-.92	.21	18.87	1	.000	.40	.26	.60

Web Appendix F: Detailed Results of Study 4

Study Design & Preregistration

2 (violation type: *targeting*, *sharing*) × 2 (privacy variability: low, high) full-factorial design. This study was preregistered on AsPredicted.org

Sample

Participants were business undergraduate students recruited from two large public universities in the US through a participant pool. We preregistered at least 800 as the sample size, but the final sample is subject to the number of students who chose to participate in the subject pool. In total, we received 952 responses from participants who were qualified for this study (N = 866 after attention check exclusions; $M_{Age} = 21.31$, $SD_{Age} = 2.55$; 5.2% female),

TABLE 2W. 21. STUDY 4: NUMBER OF PARTICIPANTS ACROSS CONDITIONS (CELL SIZES)

	Variability		
Violation Type	Low	High	**Total**
• **Targeting**	236	187	423
• **Sharing**	202	241	443
Total	**438**	**428**	**866**

Test of Discriminant Validity of Key Measures

TABLE 2W. 22. STUDY 4: PCA (WITH VARIMAX ROTATION) Results of key measures

	Component 1 (4.30; 39.11%)*	**Component 2 (2.70; 24.53%)***	**Component 3 (1.45; 13.21%)***
Ratings_1	**0.81**	-0.03	0.04
Ratings_2	**0.87**	-0.10	0.07
Ratings_3	**0.85**	-0.10	0.08
Ratings_4	**0.89**	-0.05	0.10
Ratings_5	**0.75**	0.01	0.14
Control_1	0.06	-0.13	**0.86**
Control_2	0.11	-0.19	**0.89**
Control_3	0.17	-0.24	**0.85**
Vulnerable_1	-0.06	**0.90**	-0.20
Vulnerable_2	-0.05	**0.89**	-0.23
Vulnerable_3	-0.09	**0.89**	-0.12

NOTE.—*(Component eigen value; %variance explained by component)

Manipulation Checks

TABLE 2W. 23. STUDY 4: 2 (Violation Type: Targeting, Sharing) × 2 (Privacy Variability: High, Low) Factorial ANOVAs on App Rating & Variability Check

	DV = App Rating				DV = Variability Check			
Variables	*df*	*F*	*p-value*	η_p^2	*df*	*F*	*p-value*	η_p^2
Variability	1	.06	.809	<.001	1	664.68	<.001	.435
ViolationType	1	.05	.820	<.001	1	1.47	.225	.002
Variability × ViolationType	1	.00	.959	<.001	1	4.94	.027	.006
Error	862				862			
Total	866				866			

Main Effects on Perceived Control and Vulnerability

TABLE 2W. 24. STUDY 4: 2 (Violation Type: Targeting, Sharing) × 2 (Privacy Variability: High, Low) Factorial ANOVAs on Perceived Control & Perceived Vulnerability

	DV = Perceived Control				DV = Perceived Vulnerability			
Variables	*df*	*F*	*p-value*	η_p^2	*df*	*F*	*p-value*	η_p^2
Variability	1	.02	.879	<.001	1	1.01	.314	.001
ViolationType	1	4.74	.030	.005	1	2.09	.148	.002
Variability × ViolationType	1	4.97	.026	.006	1	2.05	.152	.002
Rating	1	52.43	<.001	.057	1	19.78	<.001	.022
Error	861				861			
Total	866				866			

TABLE 2W. 25. STUDY 4: Estimated Marginal Mean of Perceived Control and Vulnerability across Conditions

Condition	DV = Perceived Control				DV = Perceived Variability			
	Mean	SE	95% CI		Mean	SE	95% CI	
			LL	UL			LL	UL
Variability = Low								
• **Targeting**	3.35	.08	3.19	3.51	4.17	.09	3.99	4.35
• **Sharing**	3.35	.09	3.17	3.54	4.17	.10	3.97	4.37
Variability = High								
• **Targeting**	3.56	.09	3.38	3.73	3.94	.10	3.75	4.13
• **Sharing**	3.17	.08	3.02	3.33	4.21	.09	4.04	4.39

NOTE—.Estimated marginal means estimated at rating = 5.38

Interaction of Violation Type and Privacy Variability on Brand Swithing

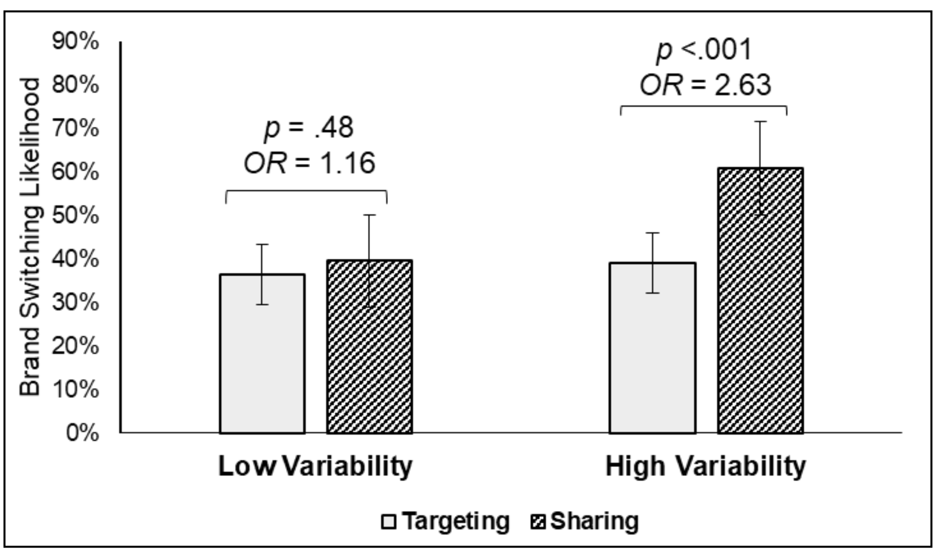
TABLE 2W. 26. STUDY 4: Wald Tests of Hierarchical Linear Logistic Regression of Violation Type and Privacy Variability on Brand Switching

Variables	*Wald* χ^2	df	p-value
Step 1: Effect of Violation Type and Variability			
• **Variability**	13.69	1	<.001
• **Violation Type**	15.68	1	<.001
Covariate: Rating	47.89	1	<.001
Step 2: Interaction Effects of Violation Type & Variability			
• **Variability**	.37	1	.544
• **ViolationType**	.50	1	.479
• **Variability × ViolationType**	7.96	1	.005
Covariate: Rating	48.29		<.001

TABLE 2W. 27. STUDY 4: Tests of Simple Effects of Violation Types at Different Levels of Privacy Variability on Brand Switching

Planned Contrasts	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Overall Main effect of Violation Types (collapsed across both conditions of variability								
• **Sharing v. targeting**	.57	.14	15.68	1	<.001	1.77	1.33	2.35
Variability = Low								
• **Sharing v. targeting**	.15	.21	.50	1	.479	1.16	.77	1.74
Variability = High								
• **Sharing v. targeting**	.97	.20	22.69	1	<.001	2.63	1.76	3.91

FIGURE 2W. 4. STUDY 4: Violation Type × Privacy Variability Interaction on Brand Switching

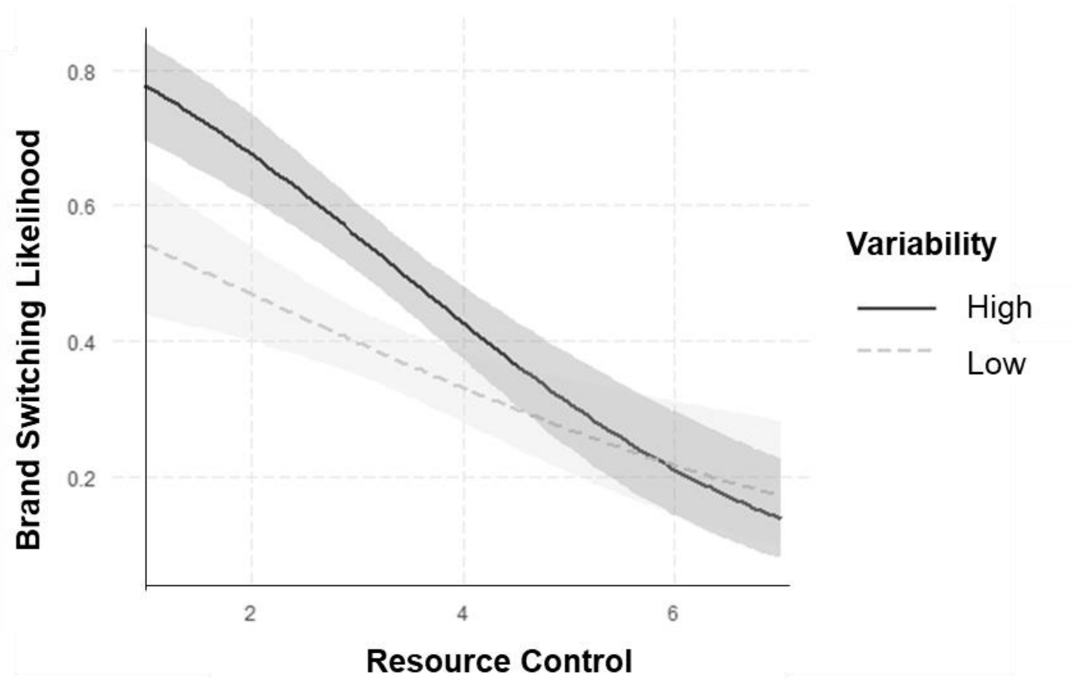


NOTE.—*p*-values represent simple contrast effects; error bars represent 95% CI; OR = odds ratio

TABLE 2W. 28. STUDY 4: Logistic Regression of Resource Control/ Vulnerability & Privacy Variability Interaction on Brand Switching

Variables	β	SE	*Wald* χ^2	df	p-value	Odds Ratio (OR)	95% C.I. OR	
							Lower	Upper
Model 1								
• **Variability**	1.46	.42	12.31	1	<.001	4.31	1.91	9.75
• **Control**	-.21	.09	5.91	1	.015	.81	.69	.96
• **Variability × Control**	-.26	.12	4.91	1	.027	.77	.61	.97
• **Rating**	-.40	.07	31.52	1	<.001	.67	.58	.77
Model 2								
• **Variability**	-.03	.51	.00	1	.956	.97	.36	2.64
• **Vulnerability**	.36	.08	19.90	1	<.001	1.44	1.23	1.69
• **Variability × Vulnerability**	.16	.12	1.98	1	.16	1.18	.94	1.48
• **Rating**	-.44	.07	36.51	1	<.001	.65	.56	.75

FIGURE 2W. 5. STUDY 4: Resource Control × Privacy Variability Interaction on Brand Switching



Moderated Mediation Analyses

TABLE 2W. 29. STUDY 4: Moderated Mediation Model 15 With Resource Control At 90% And 95% Ci

Variables	β	SE	*T*	p-value	95% C.I.	
					Lower	Upper
Model DV = Resource Control						
• **Constant**	1.90	.22	8.58	<.001	1.54	2.27
• **Violation Type**	-.19	.09	-2.22	.026	-.33	-.05
• **Rating**	.29	.04	7.23	<.001	.22	.35
Model DV = Brand Switching						
• **Constant**	2.36	.46	5.13	<.001	1.60	3.11
• **Violation Type**	.15	.21	.72	.471	-.19	.49
• **Resource Control**	-.21	.09	-2.39	.017	-.35	-.06
• **Variability**	.97	.45	2.19	.029	.24	1.71
• **Violation Type × Variability**	.72	.30	2.42	.016	.23	1.20
• **Resource Control × Variability**	-.24	.12	-1.98	.048	-.43	-.04
• **App Rating**	-.42	.07	-5.80	<.001	-.54	-.30
Indirect effects : Violation Type -> Resource control -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.04	.03			.01	.09
• **Variability = High**	.08	.04			.02	.16
Index of Moderated Mediation at 90% CI						
	Index	**BootSE**			**BootLLCI**	**BootULCI**
• **Index**	.05	.03			.002	.106
Indirect effects : Violation Type -> Resource control -> Brand Switching at different levels of variability at 95% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.04	.03			.002	.10
• **Variability = High**	.08	.04			.01	.18
Index of Moderated Mediation at 95% CI						
• **Index**	.05	.03			-.002	.12

TABLE 2W. 30. STUDY 4: Moderated Mediation Model 15 With Vulnerability At 90% Ci

Variables	β	SE	*t*	p-value	95% C.I.	
					Lower	Upper
Model DV = Vulnerability						
• **Constant**	5.11	.24	2.94	<.001	4.71	5.51
• **Violation Type**	.13	.09	1.39	.166	-.02	.29
• **Rating**	-.19	.04	-4.45	<.001	-.27	-.12
Model DV = Brand Switching						
• **Constant**	.30	.56	.53	.593	-.62	1.22
• **Violation Type**	.16	.21	.74	.461	-.19	.51
• **Vulnerability**	.36	.08	4.44	<.001	.23	.50
• **Variability**	-.43	.54	-.80	.423	-1.32	.46
• **Violation Type × Variability**	.77	.30	2.56	.010	.28	1.27
• **Vulnerability × Variability**	.16	.12	1.34	.181	-.04	.35
• **App Rating**	-.45	.07	-6.20	<.001	-.58	-.33
Indirect effects : Violation Type -> Vulnerability -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.05	.04			-.01	.11
• **Variability = High**	.07	.05			-.02	.15
Index of Moderated Mediation at 90% CI						
	Index	**BootSE**			**BootLLCI**	**BootULCI**
• **Index**	.02	.02			-.01	.06

TABLE 2W. 31. STUDY 4: Moderated Serial Mediation Model 89 With Resource Control And Vulnerability As Mediators At 90% Ci

Variables	β	SE	*t*	p-value	90% C.I.	
					Lower	Upper
Model DV = Resource Control						
• **Constant**	1.90	.22	8.58	<.001	1.54	2.27
• **Violation Type**	-.19	.09	-2.22	.026	-.33	-.05
• **App Rating**	.29	.04	7.23	<.001	.22	.35
Model DV = Vulnerability						
• **Constant**	5.90	.24	25.11	<.001	-5.52	6.29
• **Violation Type**	.05	.09	.58	.561	-.09	.20
• **Control**	-.42	.03	-12.08	<.001	-.48	-.36
• **App Rating**	-.07	.04	-1.79	.07	-.14	-.01
Model DV = Vulnerability						
• **Constant**	.43	.67	.63	.527	-.68	1.54
• **Violation Type**	.16	.21	.74	.459	-.19	.51
• **Control**	-.07	.09	-.74	.459	-.22	.08
• **Vulnerability**	.34	.09	3.96	<.001	.20	.49
• **Variability**	.67	.80	.83	.405	-.65	1.98
• **Violation Type × Variablity**	.71	.30	2.32	.020	.21	1.21
• **Control × Variability**	-.23	.13	-1.77	.076	-.44	-.02
• **Vulnerability × Variability**	.08	.12	.67	.503	-.12	.29
• **App Rating**	-.42	.08	-5.59	<.001	-.54	-.30
Indirect effects : Violation Type -> Control -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.01	.02			-.016	.05
• **Variability = High**	.06	.03			0.01	.12
Indirect effects : Violation Type -> Vulnerability -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.02	.03			-.03	.07
• **Variability = High**	.02	.04			-.04	.08
Indirect effects : Violation Type -> Control ->Vulnerability -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.028	.015			.01	.06
• **Variability = High**	.034	.018			.01	.07
Index of Moderated Mediation at 90% CI						
	Index	BootSE			BootLLCI	BootULCI
• **Index via control**	.04	.03			.004	.11
• **Index via vulnerability**	.004	.015			-.02	.03
• **Index via control -> vulnerability**	.007	.012			-.01	.03

TABLE 2W. 32. STUDY 4: Moderated Mediation Model 15 With Resource Control As Mediator And Vulnerability As Covariate At 90% And 95% Ci

Variables	β	SE	*t*	p-value	95% C.I. Lower	95% C.I. Upper
Model DV = Resource Control						
• **Constant**	1.90	.22	8.58	<.001	1.54	2.27
• **Violation Type**	-.19	.09	-2.22	.026	-.33	-.05
• **Rating**	.29	.04	7.23	<.001	.22	.35
Model DV = Brand Switching						
• **Constant**	.19	.58	.33	.743	-.94	1.32
• **Violation Type**	.16	.21	.74	.457	-.26	.58
• **Control**	-.06	.09	-.60	.546	-.23	.12
• **Variability**	1.11	.46	2.39	.017	.20	2.01
• **Violation Type × Variability**	.70	.30	2.31	.021	.11	1.30
• **Control × Variability**	-.26	.12	-2.07	.039	-.50	-.01
• **Vulnerability**	.38	.06	6.21	.000	.26	.51
• **App Rating**	-.42	.08	-5.57	.000	-.56	-.27
Indirect effects : Violation Type -> Control -> Brand Switching at different levels of variability at 95% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.01	.02			-.020	.05
• **Variability = High**	.05	.03			-.003	.12
Index of Moderated Mediation at 95% CI						
	Index	BootSE			BootLLCI	BootULCI
• **Index**	.04	.03			-.005	.10
Indirect effects : Violation Type -> Control -> Brand Switching at different levels of variability at 90% CI						
ab	**β**	**BootSE**			**BootLLCI**	**BootULCI**
• **Variability = Low**	.01	.02			-.013	.04
• **Variability = High**	.05	.03			.004	.10
Index of Moderated Mediation at 90% CI						
	Index	BootSE			BootLLCI	BootULCI
• **Index**	.04	.03			.004	.09

Web Appendix G: Detailed Results of Study 5

Pretest 1

Design. 2 (industry: search, e-commerce) within-subject design.

Sample. 300 Mturk participants (N = 252 after attention check exclusion; $M_{\text{Age}} = 37.83$, $SD_{\text{Age}} = 9.51$; 4.07 % female)

Procedure. In a random order, participants were asked to rate the search/ e-commerce industries in terms of their perceived contingency and the market dominance of Google and Amazon in their respective industries.

Generalized contingency. Please indicate the extent to which you agree or disagree with the following statements about the e-commerce/search industry (1 = Strongly disagree; 7 = Strongly agree)

- Consumers' privacy in the e-commerce/search industry depends upon their choice of brand
- Consumers in the e-commerce/ search industry are able to protect their own privacy through the brands they choose

Company dominance. In the search engine/e-commerce industry, Google/Amazon is...

- clearly not the market leader (=1) : clearly the market leader (=7)
- not at all dominant (=1) : extremely dominant (=7)

Results

TABLE 2W. 33. STUDY 5: PRETEST 1. Repeated Measure ANOVA Of Search/E-Commerce (Google/Amazon) On Perceived Contigency & Company Dominance

	DV = Generalized Contingency					DV = Perceived Dominance			
Source	*df*	*F*	*p-value*	η_p^2	**Source**	*df*	*F*	*p-value*	η_p^2
Industry (search v. e-commerce)					**Company (Google v Amazon)**				
• **Sphericity Assumed**	1	6.97	.009	.03	• Sphericity Assumed	1	1.02	.31	.004
• **Huynh-Feldt**	1	6.97	.009	.03	• Huynh-Feldt	1	1.02	.31	.004
Error					**Error**				
• **Sphericity Assumed**	251				• Sphericity Assumed	251			
• **Huynh-Feldt**	251				• Huynh-Feldt	251			

TABLE 2W. 34. STUDY 5: PRETEST 1. Estimated Marginal Means On Perceived Contigency & Company Dominance

Condition	**DV = Generalized Contingency**				**Condition**	**DV = Perceived Dominance**			
	Mean	SE	95% CI LL	95% CI UL		Mean	SE	95% CI LL	95% CI UL
Industry					**Company**				
• **E-commerce**	5.27	.08	5.11	5.44	• Amazon	6.14	.06	6.03	6.26
• **Search**	5.45	.07	5.31	5.59	• Google	6.10	.06	5.98	6.23

Pretest 2

Design. 2 (switch: yes = 1, no = 0) between-participants study

Sample. 200 Prolific participants (N = 182 after attention check exclusions, $M_{\text{Age}} = 38.64$, $SD_{\text{Age}} = 13.54$, 48.50% female)

Procedure. Participants viewed stimuli about Company A in Industry X, in which Company A was the market leader, Company A was sharing consumer data and Industry × has high privacy variability. Participants were then asked if they would continue to use Company A or switch to another company and then write a tweet (in Twitter format) about Company A. For examples of the open-ended responses, see table WG3.

Analysis. We then ran the open-response texts through LIWC Dictionary and obtained the frequency of all different pronouns used in these "tweets". We then conducted a series of one-way ANOVAs (switch: yes v. no) on these pronouns (note that there is no mention of "she" or "he" pronouns in this sample).

TABLE 2W. 35. STUDY 5: PRETEST 2. Example Tweets Provided By Participants

Example Responses
Just found my PREVIOUS clown of a company using our information to sell for their own profits gotta love just being a side hustle for a legit company lmao. #ImaData
Just found out Company A steals and sells our data... just lost one more customer. #fake
I don't trust any industry that sells and uses my information. Switch to another company that respects your privacy! We are not a product!
Just found out that Company A has been selling and sharing my personal information with other companies. Really disappointed.
Company × is a large industry comprised of many companies. Company × does not have a standard personal information policy and all company's in company × share your personal information in a different ways. You need to take this into account if this privacy concerns you.
Hey, Company A please don't share my personal information! #privacy
Great personalized product offers to connect me to what I need and want.
You screw Company A. All they do is steal our info and sell that shit for a profit. And y'all wonder why every ad is so personalized.
this company sucks
Are they selling my information? #companya #privacy
I just found out that Company A, of which I've been a longtime customer, shares my information with third-party companies. I will never shop with them again and I encourage everyone to consider doing the same. I did not agree to let them share my information.
Although I am a bit disappointed to hear that Company A shares my information to personalize ads for me, I know that this is the norm in today's world and this tactic is utilized by many corporations such as Facebook, Amazon, etc. I am not really worried about this issue.
They should not be using any of anyone's personal information in any way for other companies.

Results.

Overall, 53.85% (N_{Switch} = 98) of our sample decided to switch over staying with Company A (N_{Stay} = 84).

TABLE 2W. 36. STUDY 5: PRETEST 2. ANOVA (switch: yes v. no) on different pronouns

Source	*df*	*F*	*p-value*	η_p^2
DV = frequency of I				
Switch	1	.68	.412	.004
Error	180			
Total	182			
DV = frequency of they				
Switch	1	.25	.615	.001
Error	180			
Total	182			
DV = frequency of we				
Switch	1	1.23	.002	.054
Error	180			
Total	182			
DV = frequency of we + you				
Switch	1	4.91	.028	.027
Error	180			
Total	182			
DV = frequency of you				
Switch	1	.46	.497	.003
Error	180			
Total	182			

TABLE 2W. 37. STUDY 5: Estimated Marginal Means Of Different Pronouns

Condition	**Mean**	**SE**	**95% CI**	
			LL	UL
DV = frequency of I				
• **Stay**	6.82	.69	5.46	8.19
• **Switch**	6.05	.64	4.79	7.31
DV = frequency of they				
• **Stay**	2.49	.38	1.74	3.25
• **Switch**	2.23	.35	1.53	2.93
DV = frequency of we				
• **Stay**	.18	.29	-.40	.75
• **Switch**	1.44	.27	.91	1.97
DV = frequency of we + you				
• **Stay**	3.03	.59	1.87	4.18
• **Switch**	4.80	.54	3.73	5.87
DV = frequency of you				
• **Stay**	2.85	.54	1.78	3.92
• **Switch**	3.36	.50	2.36	4.35

Main Study

Design. 2 (contingency: high—Google, low—Amazon) × 2 (violation type: *targeting*, *sharing*) quasi experiments using tweets scraped from the Twitter API.

Sample. We used package Rtweet to collect 6,916 English tweets (N = 2,824 after removing duplicates, retweets and replies) from the Twitter Developer API for the one-month period from January 23 to February 22, 2022.

Results.

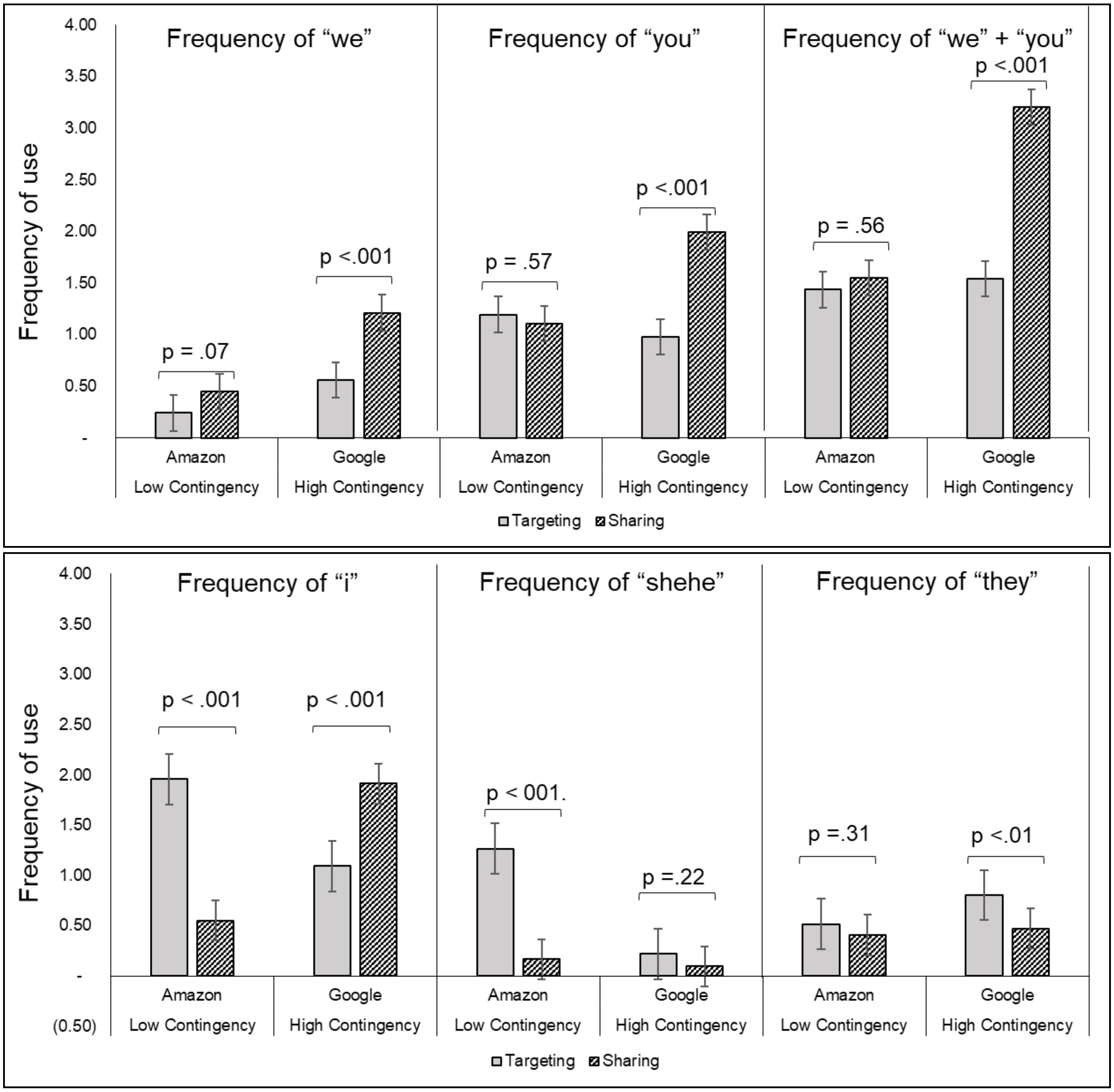
TABLE 2W. 38. STUDY 4. Examples Of Tweets Collected

	Amazon	Google
Targeting	"Amazon is LexCorp Reading that the targeted middle and low income audience pays more for Prime than if you can just pay the whole year up front. Targeted. Specifically. Yeah let's milk more from the low income. Apparently it's a working business strategy" (n= 795)	"the more Google tries to push targeted ads the more enraged i start to feel. i diverge pretty wildly with other more normie libertarian types on this in particular (many who think targeted ads are a consumer good)" (n=929)
Sharing	"Ever wonder why, when you search for a product on flipkart/amazon and the same shit starts appearing in facebook and insta. Its because apps share data with each other." (n=415)	"Google is sharing browsing history and behaviors online, not only sharing it advertisers it is also collecting your data to share with agencies like the NSA and CIA to establish patterns on you and to use against you!" (n=685)

TABLE 2W. 39. STUDY 4. Two (Contingency: Low—Amazon, High—Google) By Two (Violation Type: Targeting, Sharing) Anovas On Different Pronouns

Source	*df*	*F*	*p-value*	η_p^2
DV = frequency of we				
• Contingency	1	51.51	<.001	.018
• Violation type	1	32.43	<.001	.011
• Contingency × Violation type	1	8.69	.003	.003
• Error	2820			
• Total	2824			
DV = frequency of you				
• Contingency	1	1.22	.001	.004
• Violation type	1	19.03	<.001	.007
• Contingency × Violation type	1	27.38	<.001	.010
• Error	2820			
• Total	2824			
DV = Frequency of we + you				
• Contingency	1	45.48	<.001	.016
• Violation type	1	46.64	<.001	.016
• Contingency × Violation type	1	35.32	<.001	.012
• Error	2820			
• Total	2824			
DV = frequency of i				
• Contingency	1	4.60	.032	.002
• Violation type	1	6.46	.011	.002
• Contingency × Violation type	1	92.76	<.001	.032
• Error	2820			
• Total	2824			
DV = frequency of shehe				
• Contingency	1	5.24	<.001	.018
• Violation type	1	59.90	<.001	.021
• Contingency × Violation type	1	38.24	<.001	.013
• Error	2820			
• Total	2824			
DV = frequency of they				
• Contingency	1	6.32	.012	.002
• Violation type	1	9.81	.002	.003
• Contingency × Violation type	1	2.48	.115	.001
• Error	2820			
• Total	2824			

FIGURE 2W. 6. STUDY 5: Violation Type × Contingency Interaction On Different Pronouns



NOTE.—p-values represent planned contrast effects; error bars represent 95% CI.

References

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein (2020), "Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age," *Journal of Consumer Psychology*, 30 (4), 736–58.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016), "The Economics of Privacy," *Journal of Economic Literature*, 54 (2), 442–92.

Aguirre, Elizabeth, Dominik Mahr, Dhruv Grewal, Ko de Ruyter, and Martin Wetzels (2014), "Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness," *Journal of Retailing*, 91 (1), 34–49.

Ahmad, Irfan (2018), "How Do Social Platforms Protect Your Information? [Infographic] | Social Media Today," *Social Media Today*, September 26, https://www.socialmediatoday.com/news/how-do-social-platforms-protect-your-information-infographic/533036/.

Aiken, Leona S., Stephen G. West, and Raymond R. Reno (1991), *Multiple Regression: Testing and Interpreting Interactions*, SAGE.

Ajzen, Icek (2002), "Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior," *Journal of Applied Social Psychology*, 32 (4), 665–83.

Appel, Gil, Lauren Grewal, Rhonda Hadi, and Andrew T. Stephen (2020), "The Future of Social Media in Marketing," *Journal of the Academy of Marketing Science*, 48 (1), 79–95.

Baker, Stacey Menzel, James W. Gentry, and Terri L. Rittenburg (2005), "Building Understanding of the Domain of Consumer Vulnerability," *Journal of Macromarketing*, 25 (2), 128–39.

Bamberger, Kenneth A and Deirdre K Mulligan (2010), "Privacy on the Books and on the Ground," *Stanford Law Review*, 63, 247–316.

Benes, Ross (2018), "People Believe Ads Are Becoming More Intrusive," *eMarketer*, April 10, https://www.emarketer.com/content/people-believe-ads-are-becoming-more-intrusive.

Biron, Meghan (2019), "How Google Is Evolving for User Privacy - Think with Google," *Think with Google*, https://www.thinkwithgoogle.com/future-of-marketing/privacy-and-trust/google-user-privacy/.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Bleier, Alexander, Avi Goldfarb, and Catherine Tucker (2020), "Consumer Privacy and the Future of Data-Based Innovation and Marketing," *International Journal of Research in Marketing*, 37 (3), 466–80.

Brandeis, Samuel D. and Louis D. Warren (1890), "The Right to Privacy," *Harvard Law Review*, 4(5), 193–220.

Brough, Aaron R. and Kelly D. Martin (2020), "Critical Roles of Knowledge and Motivation in Privacy Research," *Current Opinion in Psychology*, 31, 11–15.

Cassell, Justine, David Huffaker, Dona Tversky, and Kim Ferriman (2006), "The Language of Online Leadership: Gender and Youth Engagement on the Internet.," *Developmental psychology*, 42 (3), 436.

Chung, Cindy and James W Pennebaker (2007), "The Psychological Functions of Function Words," *Social Communication*, 1, 343–59.

Confessore, Nicholas (2018), "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far - The New York Times," *The New York Times*, April 4.

Deci, Edward L and Richard M Ryan (2000), "The" What" and" Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry*, 11(4), 227–68.

Draper, Nora A and Joseph Turow (2019), "The Corporate Cultivation of Digital Resignation," *New Media & Society*, 21 (8), 1824–39.

Dweck, Carol S (2000), *Self-Theories: Their Role in Motivation, Personality, and Development*, United Kingdom: Psychology Press.

Edelman, Gilad (2022), "Proton Adds Proton Calendar, Proton Drive, and Proton VPN Encrypted Features," *Wired*, May 25.

Egan, Sean (2021), *Search Engines in the US: Competitive Landscape*.

Fast, Nathanael J., Deborah H. Gruenfeld, Niro Sivanathan, and Adam D. Galinsky (2009), "Illusory Control: A Generative Force behind Power's Far-Reaching Effects," *Psychological Science*, 20 (4), 502–8.

Foxman, Ellen R. and Paula Kilcoyne (1993), "Information Technology, Marketing Practice, and Consumer Privacy: Ethical Issues," *Journal of Public Policy & Marketing*, 12 (I), 106–19.

Friedman, Alan (2022), "Samsung's New Galaxy Television Ad is All about User Privacy," *Phone Arena*, February 13.

Grosso, Monica, Sandro Castaldo, Hua Ariel Li, and Bart Larivière (2020), "What Information Do Shoppers Share? The Effect of Personnel-, Retailer-, and Country-Trust on Willingness to Share Information," *Journal of Retailing*, 96 (4), 524–47.

Hajli, Nick and Xiaolin Lin (2016), “Exploring the Security of Information Sharing on Social Networking Sites: The Role of Perceived Control of Information,” *Journal of Business Ethics*, 133 (1), 111–23.

Hallam, Cory and Gianluca Zanella (2017), “Online Self-Disclosure: The Privacy Paradox Explained as a Temporally Discounted Balance between Concerns and Rewards,” *Computers in Human Behavior*, 68, 217–27.

Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, New York, NY, US: Guilford Press.

Henderson, Conor M., Joshua T. Beck, and Robert W. Palmatier (2011), “Review of the Theoretical Underpinnings of Loyalty Programs,” *Journal of Consumer Psychology*, 21 (3), 256–76.

Hewett, Kelly, William Rand, Roland T. Rust, and Harald J. Van Heerde (2016), “Brand Buzz in the Echoverse,” *Journal of Marketing*, 80 (3), 1–24.

Hill, Ronald Paul and Eesha Sharma (2020), “Consumer Vulnerability,” *Journal of Consumer Psychology*, 30 (3), 551–70.

Hirschman, Albert O. (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, 25, Harvard University Press.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), “Automated Text Analysis for Consumer Research,” *Journal of Consumer Research*, 44 (6), 1274–1306.

Ingram, Rhea, Steven J. Skinner, and Valerie A. Taylor (2005), “Consumers’ Evaluation of Unethical Marketing Behaviors: The Role of Customer Commitment,” *Journal of Business Ethics*, 62 (3), 237–52.

Interactive Advertising Bureau (2019), *Digital Advertising Revenues Rise To $26.2 Billion In Q3 2018, Up 22% Year-Over-Year, According To IAB*, IAB, New York, NY, US.

Jesdanun, Anick (2017), “Amazon Isn’t a ‘Traditional Monopoly’ But It Does ‘Monopolize the Attention’ of Consumers,” *Inc.com*.

Kacewicz, Ewa, James W Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C Graesser (2014), “Pronoun Use Reflects Standings in Social Hierarchies,” *Journal of Language and Social Psychology*, 33 (2), 125–43.

Kim, Tami, Kate Barasz, and Leslie K John (2018), “Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness,” *Journal of Consumer Research*, 45 (5), 906–32.

Kirkpatrick, Marshall (2010), “Facebook’s Zuckerberg Says The Age of Privacy Is Over,” *The New York Times*, January 10.

Krishna, Aradhna (2020), “Privacy Is a Concern: An Introduction to the Dialogue on Privacy,” *Journal of Consumer Psychology*, 30 (4), 733–35.

Kurland, Nancy B. and Lisa Hope Pelled (2000), “Passing the Word: Toward a Model of Gossip and Power in the Workplace,” *Academy of Management Review*, 25 (2), 428–38.

Landau, Mark J., Aaron C. Kay, and Jennifer A. Whitson (2015), “Compensatory Control and the Appeal of a Structured World,” *Psychological Bulletin*, 141 (3), 694.

Li, Han, Xin (Robert) Luo, Jie Zhang, and Heng Xu (2017), “Resolving the Privacy Paradox: Toward a Cognitive Appraisal and Emotion Approach to Online Privacy Behaviors,” *Information and Management*, 54 (8), 1012–22.

Lomas, Natasha (2019), “Google Has Quietly Added DuckDuckGo as a Search Engine Option for Chrome Users in ~60 Markets,” *Tech Crunch*, March 13.

Lynskey, Dorian (2019), “‘Alexa, Are You Invading My Privacy?’ – The Dark Side of Our Voice Assistants,” *The Guardian*, October 9.

MacInnis, Deborah J, Vicki G Morwitz, Simona Botti, Donna L Hoffman, Robert V Kozinets, Donald R Lehmann, John G Lynch Jr, and Cornelia Pechmann (2020), “Creating Boundary-Breaking, Marketing-Relevant Consumer Research,” *Journal of Marketing*, 84 (2), 1–23.

Malhotra, Naresh K., Sung S. Kim, and James Agarwal (2004), “Internet Users’ Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model,” *Information Systems Research*, 15 (4), 336–55.

Markos, Ereni, George R. Milne, and James W. Peltier (2017), “Information Sensitivity and Willingness to Provide Continua: A Comparative Privacy Study of The United States And Brazil,” *Journal of Public Policy & Marketing*, 36 (1), 79–96.

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier (2017), “Data Privacy: Effects on Customer and Firm Performance,” *Journal of Marketing*, 81 (1), 36–58.

Martin, Kelly D. and Patrick E. Murphy (2017), “The Role of Data Privacy in Marketing,” *Journal of the Academy of Marketing Science*, 45 (2), 135–55.

Martin, Kirsten (2018), “The Penalty for Privacy Violations: How Privacy Violations Impact Trust Online,” *Journal of Business Research*, 82, 103–16.

Maseeh, Haroon Iqbal, Charles Jebarajakirthy, Robin Pentecost, Denni Arli, Scott Weaven, and Md Ashaduzzaman (2021), “Privacy Concerns in E-commerce: A Multilevel Meta-analysis,” *Psychology & Marketing*, 38 (10), 1779–98.

Mathews, Lee (2021), “DuckDuckGo Is Shaking Things Up With A New Privacy-Focused Web Browser,” *Forbes*, December 21.

Matz, Sandra C., Ruth E. Appel, and Michal. Kosinski (2020), “Privacy in the Age of Psychological Targeting,” *Current Opinion in Psychology*, 31, 116–21.

Meyvis, Tom and Stijn M.J. Van Osselaer (2018), “Increasing the Power of Your Study by Increasing the Effect Size,” *Journal of Consumer Research*, 44 (5), 1157–73.

Min, Bora and Norbert Schwarz (2021), “Novelty as Opportunity and Risk: A Situated Cognition Analysis of Psychological Control and Novelty Seeking,” *Journal of Consumer Psychology*.

Nguyen, David H., Aurora Bedford, Alexander Gerard Bretana, and Gillian R. Hayes (2011), “Situating the Concern for Information Privacy through an Empirical Study of Responses to Video Recording,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3207–16.

Nicholas, Michael K., Peter H. Wilson, and Jocelyn Goyen (1992), “Comparison of Cognitive-Behavioral Group Treatment and an Alternative Non-Psychological Treatment for Chronic Low Back Pain,” *Pain*, 48 (3), 339–47.

Nivea, Richard (2019), “At F8, Zuckerberg Unveils Facebook’s New Mantra: ‘The Future Is Private,’” *CNET*, April 30.

Norberg, Patricia A., Daniel R. Horne, and David A. Horne (2007), “The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors,” *Journal of Consumer Affairs*, 41 (1), 100–126.

Norton (2020), “Protecting Your Social Media Privacy | Norton,” *Norton*, https://us.norton.com/internetsecurity-privacy-protecting-privacy-social-media.html.

Nowak, Glen J. and Joseph E. Phelps (1992), “Understanding Privacy Concerns: An Assessment of Consumers’ Information-Related Knowledge and Beliefs,” *Journal of Direct Marketing*, 6 (4), 28–39.

O’Flaherty, Kate (2018), “This Is Why People No Longer Trust Google And Facebook With Their Data,” *Forbes*, October 28.

Packard, Grant and Jonah Berger (2020), “Thinking of You: How Second-Person Pronouns Shape Cultural Success,” *Psychological Science*, 31 (4), 397–407.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), “(I’m) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm Interactions,” *Journal of Marketing Research*, 55 (4), 541–55.

Paul, Kari (2020), “They Know Us Better Than We Know Ourselves’: How Amazon Tracked My Last Two Years of Reading,” *The Guardian*, February 3.

Pew (2019), *Americans and Privacy: Concerned, Confused and Feeling Lack Of Control over Their Personal Information*, November Pew Research Center, Washington, DC.

Phelps, Joseph, Glen Nowak, and Elizabeth Ferrell (2000), "Privacy Concerns and Consumer Willingness to Provide Personal Information," *Journal of Public Policy & Marketing*, 19 (1), 27–41.

Plangger, Kirk and Matteo Montecchi (2020), "Thinking Beyond Privacy Calculus: Investigating Reactions to Customer Surveillance," *Journal of Interactive Marketing*, 50, 32–44.

Porter, Michael E (1997), "Competitive Strategy," *Measuring business excellence*.

Puntoni, Stefano, Rebecca Walker Reczek, Markus Giesler, and Simona Botti (2021), "Consumers and Artificial Intelligence: An Experiential Perspective," *Journal of Marketing*, 85 (1), 131–51.

PWC (2017), "Consumer Intelligence Series: Protect Me," (September), 4, http://www.pwc.se/sv/media/assets/consumer-intelligence-series-box-office-trends.pdf.

Reich, Brandon J. and Hong Yuan (2019), "A Shared Understanding: Redefining 'Sharing' from a Consumer Perspective," *Journal of Marketing Theory and Practice*, 27 (4), 430–44.

Romanoff, Tom (2022), "The American Innovation and Choice Online Act: What It Does and What It Means," *Bipartisan Policy*, https://bipartisanpolicy.org/explainer/s2992/.

Schiele, Kristen and Alladi Venkatesh (2016), "Regaining Control through Reclamation: How Consumption Subcultures Preserve Meaning and Group Identity after Commodification," *Consumption Markets & Culture*, 19 (5), 427–50.

Schmitt, Bernd (2019), "From Atoms to Bits and Back: A Research Curation on Digital Technology and Agenda for Future Research," *Journal of Consumer Research*, 46 (4), 825–32.

Sedek, Grzegorz, Miroslaw Kofta, and Tadeusz Tyszka (1993), "Effects of Uncontrollability on Subsequent Decision Making: Testing the Cognitive Exhaustion Hypothesis," *Journal of Personality and Social Psychology*, 65 (6), 1270–81.

Sheehan, Kim Bartel (2002), "Toward a Typology of Internet Users and Online Privacy Concerns," *The Information Society*, 18 (1), 21–32.

Sheehan, Kim Bartel and Mariea Grubbs Hoy (1999), "Using E-Mail to Survey Internet Users in the United States: Methodology and Assessment," *Journal of Computer-Mediated Communication*, 4 (3), JCMC435.

Singh, Jagdip (1991), "Industry Characteristics and Consumer Dissatisfaction," *Journal of Consumer Affairs*, 25 (1), 19–56.

Skinner, Ellen A. (1996), "A Guide to Constructs of Control," *Journal of Personality and Social Psychology*, 71 (3), 549–70.

Skinner, Ellen A. and Melanie J. Zimmer-Gembeck (2011), "Perceived Control and the Development of Coping," in *The Oxford Handbook of Stress, Health, and Coping*, ed. S. Folkman, Oxford University Press, 35–59.

Smith, H Jeff, Tamara Dinev, and Heng Xu (2011), "Information Privacy Research: An Interdisciplinary Review," *MIS Quarterly*, 35 (4), 989–1015.

Smith, H Jeff, Sandra J. Milberg, and Sandra J. Burke (1996), "Information Privacy: Measuring Individuals' Concerns about Organizational Practices," *MIS Quarterly*, 167–96.

Spiekermann, Sarah and Jana Korunovska (2017), "Towards a Value Theory for Personal Data," *Journal of Information Technology*, 32 (1), 62–84.

Summers, Christopher A., Robert W. Smith, and Rebecca Walker Reczek (2016), "An Audience of One: Behaviorally Targeted Ads as Implied Social Labels," *Journal of Consumer Research*, 43 (1), 156–78.

Sundie, Jill M., Betsy D. Gelb, and Darren Bush (2008), "Economic Reality versus Consumer Perceptions of Monopoly," *Journal of Public Policy & Marketing*, 27 (2), 178–81.

Thimm, Caja, Ute Rademacher, and Lenelis Kruse (1995), "' Power-Related Talk' Control in Verbal Interaction," *Journal of Language and Social Psychology*, 14(4), 382–407.

Thomas, Brigette (2021), "E-Commerce and Online Auctions in the US: Competitive Landscape," *IbisWorld*.

Tucker, Catherine E. (2014), "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, 51 (5), 546–62.

Vesanen, Jari and Mika Raulas (2006), "Building Bridges for Personalization: A Process Model for Marketing," *Journal of Interactive Marketing*, 20 (1), 5–20.

Walker, Kristen L. (2016), "Surrendering Information through the Looking Glass: Transparency, Trust, and Protection," *Journal of Public Policy & Marketing*, 35 (1), 144–58.

Weisz, John R. (1980), "Developmental Change in Perceived Control: Recognizing Noncontingency in the Laboratory and Perceiving It in the World.," *Developmental Psychology*, 16 (5), 385.

Weisz, John R. and Deborah J. Stipek (1982), "Competence, Contingency, and the Development of Perceived Control," *Human Development*, 25 (4), 250–81.

Wert, Sarah R and Peter Salovey (2004), "A Social Comparison Account of Gossip," *Review of General Psychology*, 8 (2), 122–37.

Whitson, Jennifer A. and Adam D. Galinsky (2008), "Lacking Control Increases Illusory Pattern Perception," *Science*, 322 (5898), 115–17.

Wuerthele, Mike (2019), "'Privacy. That's IPhone' Ad Campaign Launches, Highlights Apple's Stance on User Protection," *AppleInsider,* March 14.

Xu, Heng, Hock-Hai Teo, Bernard C. Y. Tan, and Ritu Agarwal (2012), "Effects of Individual Self-Protection, Industry Self-Regulation, and Government Regulation on Privacy Concerns: A Study of Location-Based Services," *Information Systems Research,* 23 (4), 1342–63.