

Alluring or Alarming? The Polarizing Effect of Forbidden Knowledge in Political Discourse

Personality and Social Psychology Bulletin
2026, Vol. 52(4) 743–758
© 2024 by the Society for Personality and Social Psychology, Inc
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01461672241288332
journals.sagepub.com/home/pspb



V.A. Parker¹ , E. Kehoe², J. Lees³, M. Facciani⁴,
and A.E. Wilson²

Abstract

“Forbidden knowledge” claims are central to conspiracy theories, yet they have received little systematic study. Forbidden knowledge claims imply that information is censored or suppressed. Theoretically, forbidden knowledge could be alluring or alarming, depending on alignment with recipients’ political worldviews. In three studies ($N = 2363$, two preregistered), we examined censorship claims about (conservative-aligned) controversial COVID-19 topics. In Studies 1a and 2 participants read COVID-19 claims framed as censored or not. Conservatives reported more attraction to and belief in the claims, regardless of censorship condition, while liberals showed *decreased* interest and belief when information was presented as censored. Study 1b revealed divergent interpretations of suppression motives: liberals assumed censored information was harmful or false, whereas conservatives deemed it valuable and true. In Study 2, conservatives made more critical thinking errors in a vaccine risk reasoning task when information was framed as censored. Findings reveal the polarizing effects of forbidden knowledge frames.

Keywords

censorship, misinformation, critical thinking, political polarization, conspiracy theories

Received January 29, 2024; revision accepted September 11, 2024

Introduction

The current era has been dubbed the information age (Tucci, 2023). However, along with the proliferation of information is the fast spread of *misinformation*, prompting increased attention to combatting it, sometimes including calls for its restriction (Ecker et al., 2022; Kubin et al. 2024; Pennycook & Rand, 2021). Similarly, social media has both democratized access to a public megaphone and provided new means of speech avoidance and suppression from interpersonal (blocking, unfriending, and “canceling”) to institutional (shadow-banning, content moderation, and deplatforming) forms of censorship (Ashokkumar et al., 2020; Jaidka et al., 2023; Settle, 2018). Paradoxically, society may be in a time of both great access to information and a heightened vigilance for signs of censorship and suppression in communicative environments. As many content moderation processes are not entirely transparent, it is possible for people to suspect or claim censorship without direct evidence of it (Jaidka et al., 2023; Nicholas, 2022). Indeed, while a 2020 Pew poll revealed that 73% of US adults reported believing that social media censors politically objectionable content (Vogels et al., 2020), the accuracy of this perception of censorship is not straightforward to adjudicate

(Barrett & Sims, 2021; Boone, 2023; Vogels et al., 2021). Perceptions of censorship may reflect a combination of fact (observation of actual censorship), fiction (imagined censorship when none occurred), and framing (where communicators *claim* censorship while simultaneously sharing the purportedly forbidden knowledge).

How does the perception of censorship affect people’s response to the information? On one hand, the notion that people will become more attracted to information they have

¹Kellogg School of Management, Northwestern University, Evanston, IL, USA

²Wilfrid Laurier University, Waterloo, ON, Canada

³Princeton University, Princeton, NY, USA

⁴University of Notre Dame, Notre Dame, IN, USA

J. Lees is now affiliated to Department of Human Resource Management and Organizational Behavior, University of Groningen, Groningen, The Netherlands

Corresponding Author:

V.A. Parker, Dispute Resolution Research Center, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, USA.

Email: victoria.parker@kellogg.northwestern.edu

been denied seems so evident as to be almost a truism. The allure of forbidden knowledge is a frequent theme in myth and literature from the biblical forbidden fruit to Pandora's Box. Anecdotal evidence suggests that learning about attempts to restrict or censor information can increase public attention to it, a phenomenon known as the Streisand Effect (Jansen & Martin, 2015).¹ Further, research has demonstrated that censorship can intensify the psychological appeal of information, and even increase agreement with the prohibited view. For instance, when students were told that a message they expected to hear had been censored, they became more interested in hearing it and reported more favorable attitudes toward the censored message (Worchel & Arnold, 1973; Worchel et al., 1975). At least two mechanisms may underlie the allure of forbidden knowledge. First, because people respond to scarcity by increasing the valuation of the rare commodity, restricted information may appear more exclusive and valuable, appealing to people's desire for uniqueness (Imhoff & Lambert, 2017; Worchel, 1992). Second, because the restriction of information may be experienced as a threat to personal autonomy, censorship could increase the appeal of the suppressed information via psychological reactance (Brehm, 1966; Rosenberg & Siegel, 2018; Worchel, 1992). Indeed, because suppressed information is thought to be so alluring, communicators may sometimes *claim* they have been suppressed in a bid for attention.

Experimental research systematically varying censorship (i.e., telling people information is restricted vs freely available) has received relatively little systematic research attention in social psychology since the 1970s. Recent research has focused on people's reactions to real-world censorship and efforts to circumvent government information-suppression in Turkey, Iran, and China (Behrouzian et al., 2016; Roberts, 2020). However, little is known about how people react to rhetorical *claims* of censorship. A communicator can strategically *claim* information has been suppressed, while simultaneously offering this "forbidden knowledge."

Although we argue that the psychology of censorship is under-investigated, related topics have received extensive attention. Claims of "forbidden knowledge" are a central component of conspiracy theories, which, by definition, explain real world events by attributing them to *secret* or *concealed* plots by powerful people or groups (Douglas et al., 2019). The antecedents and consequences of conspiracy theories have been well investigated (Douglas, 2021; Douglas et al., 2019; Jolley et al., 2020), but little research has focused specifically on the causal role of perceived censorship.

The classic notion of *forbidden fruit*—and the early research on censorship (Worchel & Arnold, 1973; Worchel et al., 1975)—straightforwardly suggests that censorship claims would *increase* the allure of this forbidden information, regardless of its content or source. However, the psychological literature also points to reasons why forbidden knowledge claims might not *always* be alluring. Censorship claims could be received and interpreted differently

depending on content or source—specifically, whether the information aligns or conflicts with previously held beliefs, values, or ingroup identity (Pereira et al., 2023) or comes from a source that signals those allegiances (Petrocelli, 2022). Social identity theory suggests that the groups to which people belong (including political ingroups) are a powerful source of self-worth, and as a result people tend to favor information that aligns with ingroup attitudes (Iyengar & Westwood, 2015). In the context of political identity, this dynamic can contribute to polarization, as people uncritically accept information that reinforces their pre-existing political beliefs and apply more scrutiny to outgroup-coded claims (Ditto & Lopez, 1992; Iyengar & Westwood, 2015; Kahan, 2013; Taber & Lodge, 2006). In line with these theoretical frameworks, people are more likely to believe conspiracy theories, misinformation, and political "bullshit" when it comes from a partisan ingroup (vs outgroup) member and when it aligns with political identity or worldview (Enders et al., 2023; Pereira et al., 2023; Petrocelli, 2022). In sum, extensive converging evidence suggests that (a) individuals are sensitive to the ingroup or outgroup alignment of a claim when evaluating its truth value, (b) both conservatives and liberals are susceptible to this effect (Ditto et al., 2019; Kahan, 2013; Petrocelli, 2022), and c) both motivated reasoning mechanisms and reliance on cognitive heuristics about source trustworthiness may account for the phenomenon (Kahan, 2013; Skitka & Washburn, 2016).

This broad literature on social identity, political trust, and bullshit/misinformation detection is the theoretical foundation we start from as we consider the possible impacts of politically aligned *censorship* claims through a social identity lens. This framework leads to an intriguing and novel prediction: that the effect of censorship claims will differ depending on social identity alignment. On one hand, when people learn that politically palatable information (i.e., consistent with pre-existing views, ingroup-aligned) has been censored, this forbidden knowledge claim may well heighten its intrigue (in line with the usual finding in past research; Worchel & Arnold, 1973). On the other hand, people also often assume that if information is censored, it must be false or harmful (Kubin et al., 2024). This leads to the more novel prediction that when encountering politically *unpalatable* information (i.e., inconsistent with pre-existing views, outgroup-aligned), censorship claims might not be alluring at all, but rather could be taken as a sign to proceed with caution and heightened skepticism. If this possibility is supported, it would add meaningful theoretical nuance to the existing literature on the psychology of censorship. Second, this work contributes to the social and political identity literature: if political opponents perceive censored partisan information in divergent ways (i.e., as alluring vs alarming), political polarization could be heightened compared with when that same information is freely available (Iyengar & Westwood, 2015).

To our knowledge, there is limited research on the moderators of censorship—might it sometimes *attract* and other

times *repel*? We found only one relevant case, observed under a specific set of conditions in an early study on censorship effects (Worchel et al., 1975). When students learned that they would be denied access to a censored speech, they became *more* interested in it and *more* supportive of the message, with one exception: when the suppressed speech was on a topic they opposed, *and* the censors were experts in a group they favored, their interest in the speech diminished. For instance, students who were opposed to campus policing became more interested in and supportive of the pro-police position when they learned it was censored, *except* when the censoring body was a group they liked (YM/YWCA) who were described as experts on the issue. In that condition only, students' interest in hearing the tape and their support for the argument dropped substantially.²

The Current Research

We sought to extend the literature on the psychology of censorship in several fundamental ways: (1) by examining forbidden knowledge frames (rather than actual censorship), (2) by examining political worldview as a social identity–relevant moderator of reactions to forbidden knowledge, and (3) by examining whether forbidden knowledge frames affect not only attraction but also critical thinking about new information.

First, we sought to extend existing knowledge on the psychology of censorship by examining rhetorical *claims* of censorship—which we will refer to as *forbidden knowledge (FK) framing*. Our FK operationalization is distinct from *actual* censorship where information is made unavailable (as in Worchel & Arnold, 1973; Worchel et al., 1975). FK framing makes rhetorical *claims* that information is banned, forbidden, or “*what they don't want you to know*”) while also offering access to the information alongside these claims. Framing can be a powerful tool that directs an audience to a particular conclusion (Druckman, 2001, Chong & Druckman, 2007; Nelson et al., 1997). Framing identical information as “forbidden” or not allows us to consider whether people engage with the same social identity–relevant information differently if they see it as having been suppressed or openly available. Although we are not aware of past research systematically varying the rhetorical framing of censorship, it is arguably a feature of most conspiracy theories that describe knowledge as suppressed while simultaneously offering a glimpse into that very (forbidden) knowledge.

Second, we explore how these censorship frames are received depending on their alignment with individuals' political worldview (Dharshing et al., 2017; Druckman, 2001). Integrating past research on censorship with research on politically motivated reasoning and receptivity to conspiracy theories, we reasoned that censorship frames could heighten either the allure or suspiciousness of information depending on its alignment with worldview. Because we hypothesized different reactions depending on political

alignment, we sought a politically polarized topic for which information could plausibly be framed as suppressed. Given the timing of this research (July 2021– November 2022) we focused on issues pertaining to COVID-19, which had become increasingly politicized during this period, with conservatives expressing more skepticism about the threat of COVID-19 and the recommended health measures (Calvillo et al., 2020; Conway et al., 2021; Hegland et al., 2022), more belief in COVID-19–related misinformation (Calvillo et al., 2020), and less trust in experts (Dawson et al., 2023; McLamore et al., 2022; Zhao et al., 2020). Given public debate on whether COVID-19 information was being suppressed (e.g., Clark, 2023), we suspected a censorship frame would be plausible.

In Study 1a we examine responses to headlines framed as FK or not, and we examine attraction and belief. Study 1b explores politically diverging interpretations of censorship that could explain heightened attraction versus caution. Critically, Study 2 examines headlines in a conceptual replication of S1a, then extends the examination further to consider how an FK frame affects critical thinking about vaccine risk in a task in which cursory thinking will lead to an incorrect conclusion, but careful thinking will lead to a correct one (Kahan, et al., 2017; Washburn & Skitka, 2017).

We hypothesized that people's attraction, belief, and critical thinking could be affected by censorship frames in a way that could further polarize conclusions across political worldviews. In the interest of transparency, we highlight that our original preregistered hypotheses predicted this polarized response, but we assumed it would be primarily driven by greater attraction to and credulity for censored information among politically aligned perceivers. We did not initially (S1a) predict a reverse pattern (that censorship could evoke repulsion and skepticism) among political opponents, though by S2 we noted the possibility. Nonetheless, because the censorship–repulsion pattern emerged in both studies, we opted to include theorizing that helps contextualize this newly identified (but initially unpredicted) effect of information suppression.

Transparency and Open Science

All study materials, data, and analyses syntax can be found on the OSF (https://osf.io/whvfg/?view_only=ccd9bc11f1e34a2f84526d2b5166f750). We preregistered sample size, exclusions, and key hypotheses in Studies 1a and 2: https://osf.io/64bmp/?view_only=7775aed8a5a940b3a22000202b29a554 and https://osf.io/vw2n3/?view_only=98fcad2ea88d4bca8972eb7be513a153. However, over the course of this research our focus evolved, and our analyses go beyond original predictions. We are transparent about preregistered hypotheses versus exploratory analyses and report departures from preregistration in the supplement found here: https://osf.io/ja9pu/?view_only=9b35274b0b5e493c85d1a556ee7530e2.

Table 1. Forbidden Knowledge and Control Headlines.

Label	Forbidden Knowledge	Control
Lab Leak	“THE TRUTH about the possible lab origins of the COVID-19 virus is being kept from you. Here’s the information NO ONE is allowed to talk about.”	“The possible lab origin of the COVID-19 virus is being investigated. Here’s what we know.”
Alternative Drug	“CENSORED: New paper reporting the effectiveness of an inexpensive, re-purposed drug in preventing and treating COVID-19 has been BANNED on major social media platforms. Here’s what they don’t want you to talk about.”	“New paper reporting the effectiveness of an inexpensive, re-purposed drug in the possible prevention and treatment COVID-19. Read more here.”
Vaccine Risk	“New reports about the risks of mRNA vaccines are being SUPPRESSED: Find out what they DON’T want you to talk about.”	“New reports about the possible risks of mRNA vaccines circulating: Experts weigh in.”

Study 1a

Study 1a experimentally manipulated forbidden knowledge (FK) framing by presenting participants with three COVID-19–related headlines with or without claims of censorship. Headlines focused on three real-world controversies in 2021–2022: that COVID-19 originated in a Wuhan lab, that repurposed drugs were effective in preventing illness, and that vaccines were highly risky (Sule et al., 2023).

Method

Participants

We aimed to recruit equal numbers of self-identified American liberals and conservatives from the Prolific Platform (Palan & Schitter, 2018). Participants who did not identify as liberal or conservative ($n = 44$) or failed the attention check ($n = 34$) were excluded, leaving a final sample of $N = 1000$ (see Table S0.1 for all demographics).

Statistical Power. Sensitivity analyses run in GPower with an alpha level of .05, and a minimum power of .80 (Cohen, 1992), indicated that we could detect a small effect size of $f = .09$ for a 2×2 between-subjects ANOVA.

Procedure

Participants completed the survey via Qualtrics for \$2.00 USD. After reporting their ideology, participants were randomly assigned to either the forbidden knowledge (FK) or the control condition, where they read three randomized headlines pertaining to COVID-19. The key headline message was the same across conditions, but in the FK condition these headlines were framed as censored, whereas in the control condition headlines were framed neutrally (see Table 1). Participants reported their attraction to and belief in each headline and their perceptions of censorship.

Materials

Ideology. Participants classified themselves as either “more conservative” (1), “more liberal” (2), or “both equally” (3).

Attraction. On 11-point scales where high scores reflect greater attraction, we measured interest in, the likelihood of clicking on, and willingness to share each of the headlines. We focus primarily on the aggregate score (all three items across all three headlines; $\alpha = .93$; Table 2).

Belief. Participants rated belief in each headline from 1 (*Not at all*) to 11 (*Extremely*) in the following format “The headline suggested that [the COVID-19 virus may have come from a lab.] To what extent do you believe that this claim is true?” (composite $\alpha = .77$).

Censorship. Participants indicated whether they believed the information the headline is referring to is being censored/kept from the public from 1 (*Not at all*) to 11 (*Extremely censored*; composite $\alpha = .88$).

We analyzed key dependent variables (DVs) for each headline separately and found identical patterns of results. We therefore report and interpret the three-headline composite in the text, but provide the breakdown and statistics for separate headlines in Table 2.

Results

Perceptions of Censorship

First, a 2 (ideology: liberal, conservative) $\times 2$ (condition: FK, control) ANOVA revealed that participants in the FK condition perceived the headlines as more censored than participants in the control condition, $F(1,996) = 13.32$, $p < .001$, $\eta^2 = .013$ (Table 2). However, unexpectedly, across conditions, conservatives also perceived *far* more censorship than liberals, $F(1,996) = 4475.83$, $p < .001$, $\eta^2 = .323$. Moreover, an ideology \times condition interaction revealed that the manipulation was effective for conservatives ($p < .001$, 95% CI [.54, 1.50]) but not liberals ($p = .411$, 95% CI [–.25, .61]).

Attraction to Headlines

As predicted (see Table 2), conservatives reported significantly greater attraction to these COVID-19 headlines than

Table 2. Attraction, Belief, and Censorship Perceptions by Ideology and Censorship Condition.

		Liberals			Conservatives			Overall Total			Interaction (1,996)
		FK	Control	Total	FK	Control	Total	FK	Control	Total	<i>F, p, η²</i>
		M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	
1	Attract	2.39 _a (2.17)	4.36 _b (2.70)	3.33 (2.62)	5.77 _c (3.06)	5.98 _c (2.87)	5.87 (2.97)	3.89 (3.09)	5.08 (2.89)	4.46 (3.05)	26.40, <.001, .026
	Belief	3.30 _a (2.52)	4.07 _b (2.62)	3.66 (2.59)	8.13 _c (3.03)	8.02 _c (2.84)	8.08 (2.94)	5.44 (3.66)	5.83 (3.35)	5.63 (3.52)	6.46, .011, .006
	Censor	2.92 _a (2.64)	3.05 _a (2.56)	2.98 (2.60)	6.95 _b (3.49)	6.09 _c (3.67)	6.54 (3.60)	4.71 (3.65)	4.41 (3.45)	4.56 (3.55)	6.48, .011, .007
2	Attract	2.69 _a (2.22)	4.73 _b (2.68)	3.67 (2.65)	5.76 _c (3.01)	5.55 _c (2.92)	5.66 (2.96)	4.05 (3.01)	5.10 (2.81)	4.56 (2.96)	43.50, <.001, .042
	Belief	2.70 _a (2.30)	4.34 _b (2.56)	3.49 (2.56)	6.25 _c (3.31)	6.27 _c (2.81)	6.26 (3.07)	4.28 (3.30)	5.20 (2.84)	4.72 (3.12)	21.67, <.001, .021
	Censor	2.79 _a (2.64)	2.26 _b (2.17)	2.53 (2.44)	6.73 _c (3.55)	5.33 _d (3.66)	6.06 (3.67)	4.54 (3.65)	3.63 (3.31)	4.10 (3.51)	5.12, .024, .005
3	Attract	2.34 _a (2.13)	4.27 _b (2.40)	3.26 (2.46)	5.58 _c (3.05)	5.70 _c (2.84)	5.64 (2.95)	3.78 (3.04)	4.91 (2.70)	4.32 (2.94)	30.14, <.001, .029
	Belief	3.53 _a (2.71)	4.62 _b (2.63)	4.05 (2.72)	7.51 _c (3.00)	7.43 _c (2.90)	7.47 (2.95)	5.30 (3.46)	5.87 (3.09)	5.58 (3.30)	10.49, .001, .010
	Censor	2.56 _a (2.43)	2.40 _a (2.26)	2.49 (2.35)	6.71 _b (3.49)	5.80 _c (3.76)	6.27 (3.65)	4.41 (3.60)	3.92 (3.46)	4.17 (3.54)	3.87, .050, .004
C	Attract	2.48 _a (1.93)	4.46 _b (1.97)	3.43 (2.18)	5.69 _c (2.67)	5.74 _c (2.44)	5.71 (2.56)	3.91 (2.79)	5.03 (2.28)	4.45 (2.62)	45.86, <.001, .044
	Belief	3.18 _a (1.85)	4.35 _b (1.79)	3.74 (1.91)	7.29 _c (2.54)	7.24 _c (2.14)	7.27 (2.35)	5.01 (2.99)	5.64 (2.42)	5.31 (2.75)	21.04, <.001, .021
	Censor	2.76 _a (2.22)	2.58 _a (1.93)	2.68 (2.08)	6.78 _b (3.05)	5.76 _c (3.16)	6.29 (3.14)	4.55 (3.29)	4.00 (3.00)	4.29 (3.17)	6.51, .011, .006

Note: 1 denotes the lab leak headline, 2 denotes the alternative drug headline, 3 denotes the vaccine risk headline, and C is the composite of all three. For each separate headline, main effects of condition and main effects of partisanship were all p s < .05. Subscripts denote differences at the p < .01 level.

liberals ($F(1,996) = 248.66, p < .001, \eta^2 = .200$) and were more likely to believe them to be true ($F(1,996) = 701.35, p < .001, \eta^2 = .413$). Unexpectedly, a main effect of condition indicated that overall participants were *less* attracted to ($F(1,996) = 50.61, p < .001, \eta^2 = .048$) and believed the headlines less ($F(1,996) = 17.97, p < .001, \eta^2 = .018$) in the FK than the control condition. We did not preregister a main effect hypothesis for condition because we hypothesized different patterns might emerge for conservatives and liberals. As predicted, the interaction revealed that conservatives were higher in attraction and belief than liberals across both conditions, but the effect of ideology was considerably stronger in the FK than the control condition, suggesting that partisans’ reactions to the information became more polarized when framed as “forbidden.” Descriptively, the differences between liberal and conservative attraction to the headlines were much larger (M_{diff} for attraction = 3.21; M_{diff} for belief = 4.11) in the FK condition than in the control condition (M_{diff} for attraction = 1.28; M_{diff} for belief = 2.89), in line with our speculation that censorship could heighten belief polarization.

It is plausible that conservatives would find the information more alluring in the FK condition (compared with the control) whereas liberals might show little difference or even more skepticism in the FK than control condition. Although our preregistration did not specify predictions for these simple effects, we report them. Somewhat surprisingly, given past theory and research on censorship (e.g., Worchel et al., 1975), conservatives were *not* more attracted to purportedly censored information—they were equally attracted to and accepting of the headlines across the FK and control conditions. This may in part be due to a near-ceiling effect on perceived censorship, attraction, and belief for conservatives. Instead, it was *liberals* who

were significantly *less* attracted to, and believed *less* in, the forbidden knowledge (FK)-framed headlines than the control headlines. Thus, instead of observing special allure of forbidden knowledge, we observed its mirror image, a “*nothing-to-see-here*” effect among those for whom the headlines were politically misaligned. It is worth noting, though, that conservatives perceived headlines as highly censored in both conditions, and that perceived censorship was very strongly correlated with attraction ($r = .30, p < .001$) and especially belief ($r = .60, p < .001$), even holding both political ideology and condition constant; see Table S1.5).

Study 1a Discussion

The current study extends past research on the psychological effects of censorship by (a) examining how censorship *framing* (rather than *actual* censorship, Worchel et al., 1975) affects people’s judgments, and (b) considering whether people respond differently to censorship claims when information aligns or conflicts with political worldview. Although past research focuses primarily on the allure of censored information (e.g., Behrouzian et al., 2016; Worchel et al., 1975), we argue that censorship claims could attract some people but *repel* others by signaling a need for caution or scrutiny. Consistent with this reasoning, liberals reacted to the censorship-framed (vs control) COVID-19 headlines with more disinterest and skepticism—possibly even dismissing the censorship claim itself as “bullshit” (Petrocelli, 2022). Conversely, possibly because both these COVID-19 controversies and their censorship were more discussed in conservative media (Borah, et al., 2022) and because conservatives report more experience seeing their side’s ideas censored (Barrett & Sims,

2021), conservatives viewed the headlines as highly censored, attractive, and true across both FK and control conditions, scoring much higher than liberals on all measures even in the control condition. Conservatives may have had entrenched perspectives that were more resistant to change on the basis of a headline alone.

Experimental findings were consistent with the prediction that ideological belief polarization (the degree to which liberal and conservative perspectives diverged) would be more pronounced when claims were framed as censored, though unexpectedly the “action” was due more to liberals backing away from the claims than from conservatives embracing them more. This “*nothing-to-see-here*” effect is a novel finding that extends and adds nuance and complexity to the existing literature about the psychology of forbidden knowledge—although it may *sometimes* be alluring, it may also repel. In Study 1b, we examine more directly how censorship cues may be interpreted differently depending on political worldview.

Finally, it is worth noting that although the experimental effect of the censorship condition caused more repulsion (among liberals) than attraction (among conservatives), correlationally, greater perceived censorship strongly predicted more attraction to ($r = .39$) and belief in headlines ($r = .60$) (see Table S1.4–1.7), offering some support for the hypothesis that forbidden knowledge may heighten intrigue and may even foster unwarranted credulity.

There were several features of Study 1a that warrant cautious interpretation. First, all participants saw three (either FK or control) headlines in randomized order. In the FK condition, each headline contained different phrasing including forceful claims like CENSORED and BANNED. Exposure to multiple headlines addresses generalizability; however, the repeated FK framing and incendiary tabloid-style wording might have heightened suspicion. Further, critically, because we provided only headlines, there was little *new* information to really evaluate (either skeptically or credulously). Indeed, the headlines may have been largely consistent with conservatives’ pre-existing beliefs, contributing to their near-ceiling effects across conditions.

Study 1b

In a follow-up to Study 1a we sought to directly assess how people interpreted the *meaning* of censorship claims, reasoning that social identity alignment could fuel interpretations that either attract people and foster credulity, or repel people and heighten scrutiny. A censorship claim might increase attraction if people interpret the suppressed information as a valuable scarce resource (Imhoff & Lamberty, 2017; Worchel, 1992), a threat to freedom requiring reassertion of autonomy (i.e., reactance; Brehm, 1966; Rosenberg & Siegel, 2018; Worchel, 1992), or in resistance to powerful people seeking to suppress it (as with conspiracy thinking; Douglas et al., 2019). Conversely,

a censorship claim might *repel* people if they assume that information was suppressed with good reason, such as to correct inaccuracy or to mitigate the harm it could cause (Clark et al., 2023; Kubin et al., 2024). Thus, we build on Study 1a by unpacking how partisan social identity could shape people’s *interpretations* of the meaning of censorship for COVID-19 controversies. We anticipated that liberals and conservatives would interpret the informational value of censorship in diverging ways. Based on our prior theorizing and Study 1a, we expected conservatives to interpret censorship of worldview-consistent COVID-19 claims as valuable, as autonomy threats, and as suppressed by powerful others, and liberals to interpret these censored, worldview-inconsistent claims as justifiably censored for reasons of inaccuracy or harm.

Method

Participants

We preselected for equal numbers of American liberals and conservatives from the Prolific Research Platform (see Table S0.1). We again excluded participant who identified as “both equally” ($n = 18$), leaving a final sample of $N = 390$.

Statistical Power. Given our sample size, an alpha level of .05, and a minimum power of .80 (Cohen, 1992), sensitivity analyses using GPower indicated that we could detect a small to medium effect size of $d = .29$ in this study’s independent t-tests.

Procedure

Participants completed the study online via Qualtrics for \$1.00 USD. They were asked to imagine that they had learned that some online information about three COVID-19 topics (the lab leak, alternative treatments, and vaccine risks; see Table 3 for statements) had been censored, defined as “*This means that some news and journal articles have been suppressed or retracted, and some comments on social media about the topic have been removed.*” Participants then indicated their views about what the (hypothetical) censorship would reveal about the information itself. Notably they were asked to imagine information was censored whether they believed this to be true in reality or not. They also reported whether this kind of COVID-19 information was *actually* censored.

Materials

Reasons for Censorship. Participants were asked six questions following each statement about what the censorship *tells them* about the information. Three items conveyed that censorship could signal reasons for caution (i.e., “*The information is either completely or partially untrue,*” “*If people are exposed to this information, it could cause harm to themselves or others,*” and “*The people making the decision to censor the information likely had good reason to*”). The

Table 3. COVID-19 Statements.

Topic	Statement
Lab Leak	“Imagine that you learn that some information about the origins of the COVID-19 virus, specifically, that it was due to a lab leak in Wuhan, China, has been censored.”
Alternative Drug	“Imagine that you learn that some information about the effectiveness of an alternative drug in the prevention and treatment of COVID-19 has been censored.”
Vaccine Risks	“Imagine that you learn that some information about severe and dangerous symptoms that some have experienced following the COVID-19 vaccine has been censored.”

other three items identified ways that censorship signaled a threat to autonomy or heightened value (i.e., “*The information is actually very valuable,*” “*Some people or organizations in power want to suppress the truth,*” and “*Some people or organizations are making decisions for me about what information I’m allowed to see*”).

Actual Censorship. Finally, participants reported if this kind of information was *actually* being censored, from 1 (*Not At All Censored*) to 11 (*Extremely Censored*).

Results

Reasons for Censorship

Table 4 reports the comparisons between liberals and conservatives on each of the interpretations of censorship across all three issues as well as the issue composite. Overall, liberals were significantly more likely to interpret censorship as indicating that the information was false and harmful, and viewed censorship as *for a good reason*. Conservatives were more likely to view censored information as valuable and that the censorship was a threat to their autonomy perpetrated by the powerful.

Perceptions of Actual Censorship

Participants also reported how censored they believed these claims to be in the real world. Conservatives perceived far greater actual censorship of each of the three COVID-19 topics than did liberals (see Table 5).

Study 1b Discussion

Censorship can operate as a signal of informational value, but the meaning of censorship can be interpreted in

dramatically different ways depending on a person’s priors. For COVID-19 topics, conservatives interpreted censorship as signaling that the information was valuable, and that powerful others were denying their autonomy. Liberals, on the other hand, interpreted censorship as indicating that the information was inaccurate or harmful, and thus must have been done for good reason. Thus, forbidden knowledge frames may repel or attract based on fundamentally different views about what the censorship means, and may be a key contributor to rising levels of polarization as partisans immerse themselves in increasingly separate informational ecosystems (Wilson et al., 2020) with different claims of forbidden knowledge.

However, it is important to recall that these interpretations of censorship pertained specifically to COVID-19 headlines that were aligned with conservative worldviews. We do not claim on this basis that liberals would *always* justify censorship and conservatives would *always* interpret censored information as valuable and an autonomy threat. Instead, we speculate that if partisans perceived that *liberal*-aligned information was censored, liberals and conservatives might reverse in their interpretations of censorship in line with research showing symmetry in political bias (Ditto et al., 2019; Petrocelli, 2022).

Study 2

In Study 2, we built on Study 1a, attempting to address its limitations by (a) focusing on a single issue (COVID-19 vaccine risks) rather than presenting repeated headlines, (b) creating more visually realistic headlines for heightened engagement, and (c) including fabricated vaccine risk information in an article following the headline, so that participants would have *new* information to evaluate regardless of their pre-existing vaccine views. By including both a headline *and* an article (both framed neutrally or as censored), we first conceptually replicated Study 1 by assessing attraction and belief immediately following the headline, then extended Study 1 by examining participants’ critical thinking about the new information in the article. When reading an isolated headline, there is no new information to critically analyze; thus people may rely heavily on their prior assumptions and political worldview. The fake article, in contrast, implied evidence of new vaccine risks (claiming, in the FK condition, that this information had been suppressed). Yet, careful examination of the data presented in the article revealed no evidence of heightened vaccine risk. We adapted a method from Kahan et al. (2017) in which a table of data is presented such that a cursory look (at numerators) would support one conclusion, but a careful examination (considering denominators and, thereby, proportions) supports another conclusion. Kahan et al. (2017) found that people were more likely to critically reason to the correct conclusion when it supported their worldview—for instance, if a cursory look at evidence

Table 4. Comparisons of Liberal and Conservative Agreement with Pro- and Anti-Censorship Items.

		Liberal	Conservative	t	df	p	95% CI	d
		M(SD)	M(SD)					
Lab Leak	Likely false	4.86 (1.63)	3.36(1.88)	-8.26	348.02	<.001	[-1.85, -1.14]	-.86
	Likely harmful	4.62 (1.66)	3.43(2.10)	-6.11	329.80	<.001	[-1.58, -.81]	-.85
	Good reason	4.63 (1.65)	3.05(1.93)	-8.61	346.56	<.001	[-1.95, -1.23]	-.92
	Valuable	3.90 (1.85)	5.47(1.57)	8.93	384	<.001	[1.23, 1.92]	.91
	Powerful censor	4.44 (2.06)	6.14(1.26)	9.94	353.82	<.001	[1.36, 2.04]	.98
	Decisions	5.32 (1.65)	6.35(1.12)	7.25	368.02	<.001	[.75, 1.31]	.72
Alt Drug	Likely false	4.93 (1.96)	3.61(1.87)	-7.43	341.84	<.001	[-1.66, -.97]	-.77
	Likely harmful	4.82 (1.71)	3.35(1.82)	-8.16	363.14	<.001	[-1.83, -1.12]	-.84
	Good reason	4.58 (1.75)	3.15(1.82)	-7.83	385	<.001	[-1.78, -1.08]	-.80
	Valuable	3.74 (1.81)	5.32(1.56)	9.08	385	<.001	[1.23, 1.92]	.93
	Powerful suppress	4.07 (2.01)	5.92(1.33)	10.81	367.16	<.001	[1.51, 2.19]	1.07
	Decisions	5.22 (1.69)	6.30 (.96)	7.84	342.82	<.001	[.81, 1.35]	.77
Vax Symptoms	Likely false	4.70 (1.64)	3.64(1.97)	-5.69	341.58	<.001	[-1.43, -.70]	-.59
	Likely harmful	4.52 (1.73)	3.34(1.94)	-6.32	357.87	<.001	[-1.55, -.81]	-.65
	Good reason	4.53 (1.71)	3.00(1.85)	-7.44	383	<.001	[-1.71, -.99]	-.76
	Valuable	4.13 (1.81)	5.36(1.58)	7.07	383	<.001	[.89, 1.58]	.72
	Powerful suppress	4.36 (1.97)	6.02(1.29)	9.90	360.21	<.001	[1.33, 1.99]	.98
	Decisions	5.36 (1.63)	6.23(1.15)	6.06	372.25	<.001	[.58, 1.14]	.60
Overall	Likely false ($\alpha = .86$)	4.83 (1.37)	3.54 (1.68)	-8.19	335.42	<.001	[-1.60, -.98]	-.85
	Likely harmful ($\alpha = .85$)	4.66 (1.43)	3.37 (1.71)	-7.90	341.27	<.001	[-1.60, -.96]	-.82
	Good reason ($\alpha = .87$)	4.52 (1.46)	3.07 (1.66)	-9.06	351.56	<.001	[-1.77, -1.14]	-.94
	Valuable ($\alpha = .88$)	3.92 (1.59)	5.38 (1.37)	9.70	384.44	<.001	[1.16, 1.76]	.98
	Powerful suppress ($\alpha = .92$)	4.30 (1.86)	6.03 (1.12)	11.31	352.43	<.001	[1.43, 2.04]	1.11
	Decisions ($\alpha = .92$)	5.30 (1.55)	6.29 (.93)	7.76	352.40	<.001	[.74, 1.24]	.76

Note: Where Levene's tests were significant, results are reported wherein equal variances are not assumed.

Table 5. Comparisons Between Liberal and Conservative Ratings of Actual Censorship.

	Liberal	Conservative	t	df	p	95% CI	d
	M(SD)	M(SD)					
Lab Leak	3.60 (2.76)	7.80 (3.00)	14.27	381	<.001	[3.62, 4.78]	1.46
Alternative Drug	3.04 (2.54)	7.29 (3.36)	13.77	321.82	<.001	[3.64, 4.85]	1.44
Vaccine Risks	3.09 (2.66)	7.72 (3.36)	14.77	331.14	<.001	[4.01, 5.25]	1.54

Note: Where Levene's tests were significant, results are reported wherein equal variances are not assumed.

contradicted gun control policy but careful consideration of evidence supported it, pro-gun-control liberals would “do the math” more effectively than conservatives, a pattern that reversed when the data showed the opposite pattern (see also Washburn & Skitka, 2017). We adopted this method, however, to test a novel prediction that goes beyond the motivated reasoning account to examine whether *censorship* claims themselves could dampen critical scrutiny of worldview-consistent information. We hypothesized that conservatives would be especially likely to draw the worldview-consistent but *incorrect* conclusion from the data when the evidence was framed as censored (FK condition). Note that this prediction goes beyond past research (Kahan et al., 2017): although it is possible that conservatives will engage in less reflection because the cursory conclusion is consistent with their

ingroup's COVID-19 concerns, our critical prediction is that conservatives (more than liberals) will be especially susceptible to this kind of dampened critical thinking *following a censorship claim*.

Method

Participants

We recruited American adults from the Cloud Research Platform (see Table S0.1), aiming for equal samples of liberals and conservatives. Participants who were neither liberal nor conservative ($n = 62$) or who failed attention checks ($n = 175$) were excluded, leaving a final sample of $N = 973$ (see preregistration: https://osf.io/vw2n3?view_only=98fcd2ea88d4bca8972eb7be513a153).

Statistical Power. Our key hypothesis tests included two-way interactions between political orientation (liberal or conservative) and condition (FK or control). For an alpha level of .05, and a minimum power of .80 (Cohen, 1992), sensitivity analyses run in GPower indicated that we could detect a small effect size of $f = .10$ for the interaction and a critical X^2 of 5.99 for nonparametric tests.

Procedure

Participants completed the survey via Qualtrics for \$3.00 USD. They were randomly assigned to the forbidden knowledge (FK) or control condition and shown a visually realistic news headline about COVID-19 vaccine risks (described as censored or not, see Figure 1) and asked to report their attraction to and belief in the information (see materials: https://osf.io/whvfg/?view_only=ccd9bc11f1e34a2f84526d2b5166f750). Participants then read a news article referencing a fictitious clinical trial tracking possible severe symptoms of the COVID-19 vaccine. This article presented data that participants were subsequently asked to interpret (quantitative reasoning task modified from Kahan et al., 2017). Participants then completed the remaining dependent variables including censorship perception.

Materials

Headlines and Articles. The headlines (see Figure 1) and articles were fabricated by the researchers. The article ostensibly described a new preprint about COVID-19 vaccine risks, and included three key features to create the intended psychological conditions. First, the information contained in the headline and article were identical across FK and control conditions, except for key phrases inserted into the FK condition suggesting attempts to suppress the information (e.g., “Reporters were able to capture the data reported in the preprint before it was suddenly taken down just days later, without any explanation;” see https://osf.io/bakzp?view_only=ccd9bc11f1e34a2f84526d2b5166f750 for materials). Second, the article implied that this new research may reveal serious vaccine risks but did not offer a concrete interpretation of the data. Third, the actual data presented did not indicate an elevated risk of symptoms among those who received the COVID-19 vaccine (vs unvaccinated) but a cursory examination suggested heightened risk. Adapting the critical reasoning task from Kahan et al. (2017), participants saw a 2 × 2 table showing vaccination status (yes/no) and presence (yes/no) of severe symptoms such as blood clots, miscarriage, and death (see Table 6). Critically, the sample size of the vaccinated group was much larger (>5×) than the unvaccinated group. As a result, the absolute number of vaccinated people with severe symptoms was higher than for unvaccinated people. However, the proportion of symptoms across vaccinated and unvaccinated groups are nearly identical (9.15% in the

vaccinated group and 9.25% in the unvaccinated group). Therefore, an uncritical examination of the table (focusing on the numerator alone) could lead someone to conclude there are more cases, and thus higher risk, in the vaccinated group, but a careful examination of proportions leads to the conclusion that vaccines do not increase risk.

Measures. The ideology and attraction items ($\alpha = .91$) were identical to those in Study 1a. The belief item, assessed after exposure to the headline only, read: “The headline suggested that the COVID-19 vaccine may have severe side effects. To what extent do you believe that this claim is true?” The censorship item was measured at the end of key DVs after both the headline and article were presented, and read “Do you believe the information the headline/article is referring to is being censored/kept from the public?”

Quantitative Reasoning Test. After reading the article, participants completed the new measure of critical thinking. After examining the table of data, they were asked which conclusion was supported: (a) that vaccinated individuals are more likely to experience the severe symptoms listed (incorrect response), (b) that vaccinated people were less likely to have symptoms, or (c) that they were no more or less likely to have symptoms. Note that we considered the latter two options to both be correct responses. We designed the proportions to be as equivalent as possible reflecting no meaningful difference in risk levels—thus (c) was initially the response we intended as correct. However, on further consideration we recognized that because proportionally, the unvaccinated group was .1% more at risk, answer (b) is also technically correct. For simplicity and because we saw no meaningful theoretical distinction between response b and c, main analyses were conducted with a dichotomous (correct/incorrect) dependent variable. However results remain consistent when all three options are considered separately (see Table S2.10 & S2.11 in supplement).

Results

Perceived Censorship

A 2 (Ideology: liberal or conservative) × 2 (Condition: FK or control) ANOVA (see Table 7) indicated that FK participants perceived more censorship than controls, $F(1, 969) = 46.72, p < .001, \eta^2 = .05$, and conservatives saw the article as much more censored than liberals, $F(1, 969) = 319.68, p < .001, \eta^2 = .25$. Further, a significant ideology × condition interaction emerged (Table 7). Simple effects indicate that the FK manipulation (vs. control) successfully increased perceived censorship for both liberals and conservatives (improving on Study 1a in which liberals’ censorship perceptions did not increase), however the effect of censorship condition was greater for conservatives than liberals.



Figure 1. Visual of the forbidden knowledge (FK) and control headline conditions, Study 2.

Table 6. Experimental Stimuli: Fictitious Results of COVID-19 Vaccine Study.

	Did not experience severe symptoms	Did experience severe symptoms
Individuals that <i>did</i> receive the COVID-19 vaccine	278	28
Individuals that <i>did not</i> receive the COVID-19 vaccine	49	5

Note: The following text was presented underneath the table in the article: "Vaccination and non-vaccination group sizes differ because of vaccination rates in the population at the time of data collection. Differential respondent completion rates also contributed to sample characteristics but does not prevent assessment of results."

Attraction to and Belief in the Headline

As hypothesized, a 2 (Ideology) \times 2 (Condition) ANOVA revealed that conservatives were more attracted to ($F(1, 969) = 229.02, p < .001, \eta^2 = .191$), and had greater belief in, the COVID-19 headline ($F(1, 969) = 395.21, p < .001, \eta^2 = .290$) than liberals (see Table 7). As in Study 1a, an (unpredicted) main effect of condition indicated that people were overall *less* attracted to forbidden knowledge than neutral headlines ($F(1, 969) = 6.33, p = .012, \eta^2 = .006$) but was qualified by the predicted ideology \times condition interaction revealing that conservatives were more attracted to the headline than liberals, and this divergence was larger in the FK condition ($M_{diff} = 3.36$) than the control condition ($M_{diff} =$

2.40). Although we predicted that conservatives would be more attracted to the headline in the FK than control condition, we instead replicated the Study 1a pattern in which conservatives were highly attracted to the headline in both conditions, and *liberals* were again significantly *less* attracted to the headline in the FK than control condition, avoiding worldview-inconsistent information ("*nothing-to-see-here*") especially when claims were framed as censored. In Study 2 there was no main effect for belief ($F(1, 969) = .06, p = .813, \eta^2 = .000$), and although the interaction showed a similar pattern to attraction, it did not reach significance.

Critical Thinking

Next, we examined the central measure of critical thinking about vaccine risk: whether people drew the correct (no increased risk) or incorrect (increased risk) conclusion from the 2 \times 2 table of supposed vaccine symptom results. A Chi-square test of independence indicated that, as expected, conservatives were more likely to come to the incorrect conclusion if the information about the COVID-19 vaccine was framed as "forbidden" and suppressed than when the same information was framed as freely available, $\chi^2(1, N = 973) = 6.25, p = .012$. Conservatives were significantly more likely to incorrectly conclude that vaccinated individuals were more at risk of severe side effects when in the FK condition (57% incorrect) than when in the control condition (41% incorrect; see Table 8 for counts and contrasts); liberals did not show these differences (40% incorrect in the FK

Table 7. Attraction to and Belief in Headlines, and Perceived Censorship by Ideology and Condition.

	Liberals			Conservatives			Overall Total			Interaction (1,969)
	FK	Control	Total	FK	Control	Total	FK	Control	Total	<i>F, p, η²</i>
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	
Attract	2.09 _a (2.74)	3.05 _b (3.03)	2.60 (2.93)	5.45 _c (3.13)	5.45 _c (2.89)	5.45 (3.01)	3.66 (3.37)	4.11 (3.19)	3.89 (3.29)	6.41, .011, .007
Belief	2.30 _a (2.69)	2.56 _a (2.78)	2.44 (2.73)	6.39 _b (3.35)	6.03 _b (3.02)	6.21 (3.19)	4.20 (3.64)	4.09 (3.36)	4.14 (3.50)	2.65, .104, .003
Censor	2.67 _a (2.95)	1.81 _b (2.53)	2.22 (2.77)	6.59 _c (3.23)	4.80 _d (3.35)	5.70 (3.40)	4.50 (3.65)	3.13 (3.27)	3.79 (3.52)	5.70, .017, .006

Note: Subscripts denote differences both within participant (differences between condition for liberals and conservatives), and between participant (differences between liberals and conservatives within each condition) at the $p < .05$ level. Bonferroni corrections used for multiple tests.

Table 8. Partisanship and Condition on Critical Thinking.

		Forbidden Knowledge	Control
Conservatives	Incorrect Conclusion	125 _a	91 _b
	Correct Conclusion	95 _a	129 _b
Liberals	Incorrect Conclusion	100 _a	108 _a
	Correct Conclusion	152 _a	173 _a

Note: Table displays counts. Differing subscripts (read within each row) denote differences at the $p < .01$ level.

condition, 38% in the control). Further, conservatives were more likely than liberals to come to the incorrect conclusion in the FK condition ($p < .001$), but not in the control condition ($p = .521$) (see Table S2.9 in supplement).³

Study 2 Discussion

Study 2 extended Study 1a in several ways. First, we conceptually replicated Study 1a by again manipulating FK framing, but improved study realism by presenting a single, realistic headline and additional information to evaluate in an article. In Study 2, the FK framing successfully increased perceived censorship for both conservatives *and* liberals (in Study 1a it only affected conservatives), suggesting the increased realism may have increased overall impact of the manipulation; however, conservatives across both FK and control conditions still saw the headline and article as significantly more censored than liberals.

As in Study 1a, we first obtained reactions to the headline alone. Departing from past research (e.g., Worchel et al., 1975), the forbidden knowledge frame did *not* make headlines more alluring overall—conservatives saw headlines as equally attractive and believable across conditions, whereas liberals were significantly *less* attracted to the headline when presented with FK framing, suggesting that censorship gave them reason to avoid the worldview-inconsistent information. This “*nothing-to-see-here*” effect appears to be a novel contribution to knowledge about the effects of censorship and suggests a way that partisans may polarize under conditions where speech is at least perceived to be

suppressed. Just as some people find forbidden knowledge alluring, others are repelled. In Study 1a, FK framing also reduced liberals’ belief in the headline, an effect not replicated here. In Study 2, belief in the headline alone was high for conservatives across both conditions and substantially lower for liberals. This may reflect the increasing polarization on this topic of vaccines by the time this study was run in March 2022 (e.g., Hegland et al., 2022). When judging the headline alone, people’s reactions may have been strongly influenced by pre-existing assumptions informed by political worldview.

However, Study 2 went beyond obtaining ratings of headlines; participants were also given *new* information to evaluate in an article. The information was misleading in that the article implied vaccine risk, and a quick perusal of the data could lead to a (false) conclusion of heightened risk, while careful evaluation of the data (calculating proportions) revealed no added risk. As expected, conservatives led to believe a COVID-19 vaccine risk preprint had been mysteriously taken down (implying suppression of findings) were more likely to incorrectly conclude that the study revealed vaccine risks (the conclusion suggested by a cursory examination of the data), even though careful examination of evidence contradicted this perspective. A majority of conservatives drew the wrong conclusion when the information was framed as forbidden knowledge. In contrast, conservatives in the control condition, and liberals in both conditions, were less likely to come to the incorrect conclusion. This suggests that when information is appealing to one’s worldview, a rhetorical claim of censorship (even when the information is still readily available) may be enough to encourage uncritical acceptance of “forbidden knowledge” claims and hamper critical thinking about readily available evidence to the contrary.

Although in all conditions except the conservative FK condition, data are consistent with a more reflective evaluation of the article (concluding “no additional risk” from the data, in spite of the article strongly implying vaccine risk), it is important to acknowledge that liberals selecting “the no additional vaccine risk” options could have simply been relying on their own pre-existing beliefs that vaccines

are safe. Because it happens to be that for liberals the correct answer is aligned with their preferred answer, we can't be confident they actually reflectively evaluated the data (i.e., calculating proportions to arrive at the correct conclusion). However, it is quite notable that conservatives in the *control* condition came to the correct conclusion as often as liberals in both conditions. This finding does suggest stronger evidence that conservatives evaluated the data more critically in the control condition than the FK condition, since the correct conclusion goes against their worldview-consistent belief. In short, this study provides especially compelling evidence that when people believe that worldview-consistent information has been censored (in this context, conservatives), they may be less likely to think critically about those claims.

General Discussion

In the current digital age, information—both true and misleading—is everywhere, and both real and alleged attempts at suppression are commonplace. This research updates and extends classic work on the psychology of censorship (e.g., Worchel, 1992) by considering its effects through the lens of social identity theory (Iyengar & Westwood, 2015; Kahan, 2013). We contribute new knowledge to the literature on psychological censorship by illuminating how censorship claims may be received and interpreted differently when the claims are congruent or incongruent with pre-existing views and identities (Iyengar & Westwood, 2015; Kahan, 2013). Despite common assumptions about the allure of the forbidden, the patterns we observed were more complex. Although people who perceived greater censorship did indeed report more attraction to and belief in controversial COVID-19 claims, it isn't simply the case that claiming censorship always increases allure. Instead, when the informational claim was inconsistent with partisans' political worldview, rhetorical claims of censorship instead reduced attraction to the information (i.e., interest, desire to click and to share the headline) and heightened skepticism. Because we focused on COVID-19 controversies more aligned with conservative worldview, we expected forbidden knowledge in this domain to be especially alluring to conservatives and for censorship claims to heighten the polarization between conservative and liberal conclusions about these claims. We did observe this polarization, though unexpectedly it was driven by liberals' *repulsion* from the censored COVID-19 claims—a "*nothing-to-see-here*" effect. We saw less movement among conservatives' responses to headlines, likely because they were already near-ceiling on how censored, attractive, and believable they saw them to be. Critically, however, conservatives *were* affected by forbidden knowledge claims when encountering new, misleading but worldview-consistent information (Study 2). In this study (adapted from the judgment task in Kahan et al., 2017), conservatives who read an article claiming censorship were less likely to

critically think about the vaccine risk data it presented (which in fact showed no added risk) compared with conservatives who read the identical article with no forbidden knowledge claim.

Study 1b reveals how these polarized responses to censorship may occur because censorship can be interpreted in varied ways. Some people may attribute prosocial motives to censorship, assuming that if information is censored, it must be false or harmful (Clark et al., 2023; Kubin et al., 2024). Others may believe that censorship means that powerful others are attempting to restrict autonomy and withhold valuable knowledge. In the context of COVID-19 claims, liberals interpreted censorship in the former manner and conservatives in the latter way. This partisan divergence in responses to censored information may contribute to rising polarization (e.g., Wilson et al., 2020): while one political side may too-credulously fixate on restricted information, the other side may too-dismissively treat all such claims with skepticism. This phenomenon may offer a snapshot of information seeking mechanisms that exacerbate polarization over time as people approach, avoid, and trust different narratives from increasingly different informational ecosystems.

What might these findings mean in the broader sphere of social discourse, media, and alternative news producers? As people encounter perceived misinformation, they may call for censorship, perhaps regarding these demands to be prosocial (Olshansky & Landrum, 2020). Yet if others come to believe that information consistent with their worldview is restricted (even if, objectively, it isn't), they may be motivated to resist the perceived censorship (Behrouzian et al., 2016) by seeking out alternative information sources (often of dubious quality) that purport to have access to the forbidden truth. These "alternative" sources may include explicit conspiracy theorists like Q-Anon, and also many others in the alternative media space (of varying quality and extremity) who may deliberately attract attention using a forbidden knowledge rhetorical frame. These individuals have thriving alternative media platforms on YouTube, Substack, or podcasts built partly on offers of forbidden knowledge (Weiss, 2018). Importantly, we do not make any normative claim about whether such alternative information should be avoided. Indeed, sometimes counter-mainstream ideas *should* be considered, albeit with the same critical lens as any other information. However, if people evaluate a claim simply by its purported status as "censored," they may give these ideas more (or less) attention and critical scrutiny than warranted.

The proliferation of misinformation was rampant during the COVID-19 pandemic, with the WHO describing its spread through social media as an "infodemic" (Zarocostas, 2020). Yet, combatting misinformation through censorship could backfire in meaningful ways if those most receptive to misleading information become more susceptible to it. Some researchers have argued that on balance, removal of extreme sources of false information is warranted and effective by

some metrics: deplatforming Alex Jones from Twitter, for example, reduced the number of people talking about him, and decreased the levels of toxicity among his supporters (Jhaver et al., 2021). Others, however, argue that deplatforming has unintended consequences, such as pushing those with censored views to increasingly isolated parts of the internet where congregation of like-minded extremists can foster radicalization (Rogers, 2020). Moreover, perceiving censorship may undermine public trust, with profound implications for matters from public health to election integrity (British Royal Society, 2022).

Limitations and Future Directions

Although this research opens new avenues of inquiry on the polarizing effects of forbidden knowledge, the current paper has limitations to be addressed in future research. First, we focused on a set of important real-world issues around COVID-19 where misinformation risks were high and censorship was a hot topic in social discourse. This focus generated timely and ecologically valid knowledge applicable both to the pandemic and to future public health and safety crises. However, the focus on COVID-19 limits generalizability, and thus future research should address this limitation by examining other social issues. Further, because of the nature of COVID-19 political polarization, these issues were most consistent with conservative worldviews (Hegland et al., 2022). Future work should examine parallel issues that could appeal more to liberals or other social identities. Finally, because we studied a hot contemporary issue, some attitude differences may have been near-ceiling at the outset, limiting additional movement on already-familiar topics. Although it is valuable to study real-world issues on matters of public importance, it would also be helpful to examine more neutral issues on which opinions are not yet entrenched.

In our studies, conservatives consistently perceived higher censorship (across conditions) and issue interest while liberals showed marked *disinterest* in censored information. We do not, however, claim that liberals will always avoid allegedly censored information or that conservatives will always be prone to more errors of critical reasoning about forbidden knowledge. Indeed, in line with past research finding symmetrical cross-partisan effects of bias, misinformation, and bullshit detection (Ditto et al., 2019; Kahan, 2013; Petrocelli, 2022), we strongly suspect that liberals would also be susceptible both to the allure of forbidden knowledge and its disruption to critical thinking when they encounter worldview-congruent forbidden knowledge.

Although we expect that people of any ideology could at times be vulnerable to forbidden knowledge, we cannot rule out the possibility that some people are especially prone to it. For instance, in contrast with the partisan symmetry perspective (Ditto et al., 2019; Kahan, 2013; Petrocelli, 2022), some past research suggests asymmetry—that conservatives may be more biased, possibly due

to individual differences such as lower cognitive reflection or need for closure (Deppe et al., 2015; Skitka & Washburn, 2016). Although future research is needed to examine the limits of symmetry, we offer preliminary evidence in the supplement that individual differences in cognitive reflection (sometimes lower in conservatives) does not account for the patterns we observe (see page S6 of supplement). Even individual differences in media consumption and institutional trust could underlie reactivity to censorship claims. For instance, conservatives report having experienced or witnessed more censorship on social media (Barrett & Sims, 2021). Exploratory analyses on media consumption and individual differences are reported in the supplemental in the interest of generating future research (see Tables S1.4–1.7 and S2.4–2.7).

Finally, in the present work, we have focused on one facet of forbidden knowledge: censorship frames. Future research should examine other ways knowledge may come to be seen as “forbidden.” For instance, information could also be suppressed more subtly by emphasizing reputational costs of discussing controversial or taboo information openly, resulting in self-censorship.

Conclusion

While myth and literature often describe the forbidden as impossible to resist, people may not *all* be destined to suffer Pandora’s classic fate when faced with forbidden knowledge. We point to Pandora’s paradox instead: censorship may have the capacity to be alluring *or* repellent, to trigger credulity or intensify scrutiny. To the extent that censorship claims lead partisans to seek, avoid, and evaluate information along drastically different pathways, information suppression could play a pivotal role in political polarization and the fragmenting of epistemic ecosystems. Although censorship is sometimes thought of as a tool to fight misinformation, greater consideration should be given to this double-edged sword.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research project was funded by the Social Sciences and Humanities Research Council (SSHRC) to the first and last author (grant number 435-2019-1034), the Canadian Institute for Advanced Research (CIFAR) to the last author, and from Meta’s Misinformation and Polarization research grant to the first and last author.

ORCID iD

V.A. Parker  <https://orcid.org/0000-0003-0614-0162>

Supplemental material

Supplemental material is available online with this article.

Notes

1. The term *Streisand effect* was coined after Barbara Streisand sought to suppress a photo of her mansion in the California Coastal Records Project. The image was downloaded six times prior to the lawsuit and 420,000 times in the month after the suppression attempt was made public.
2. It is worth noting that these very interesting patterns reflect a complex three-way interaction obtained in a study with a sample size typical of the 1970s, considered severely underpowered by today's standards. We consider this finding conceptually intriguing but far from conclusive.
3. As noted previously we also report the results of this analysis breaking the "correct conclusion" variable into two separate response options (see supplementary materials, Table S2.10 and S2.11); the conclusions remain identical.

References

- Ashokkumar, A., Talafar, S., Fraser, W.T., Landabur, R., Buhrmester, M., Gómez, Á., Paredes, B., & Swann, W.B. (2020). Censoring political opposition online: Who does it and why. *Journal of Experimental Social Psychology, 91*, 104031. <https://doi.org/10.1016/j.jesp.2020.104031>
- Barrett, P.M., & Sims, J.G. (2021). False accusation: The unfounded claim that social media companies censor conservatives. *New York University, Stern School of Business, Center for Business and Human Rights*. Retrieved from <https://perma.cc/X5A8-HQLB>
- Behrouzian, G., Nisbet, E.C., Dal, A., & Çarkoğlu, A. (2016). Resisting Censorship: How Citizens Navigate Closed Media Environments. *International Journal of Communication, 10*, 4345–4367.
- Boone, R. (2023, March 15). *Experts say attacks on free speech are rising across the U.S.* PBS. <https://www.pbs.org/newshour/politics/experts-say-attacks-on-free-speech-are-rising-across-the-us>
- Borah, P., Lorenzano, K., Vishnevskaya, A., & Austin, E. (2022). Conservative media use and Covid-19 related behavior: The moderating role of media literacy variables. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph18126318>
- Brehm, J.W. (1966). *A theory of psychological reactance*. Academic Press.
- British Royal Society. (2022). *The online information environment: Understanding how the internet shapes people's engagement with scientific information*. <https://royalsociety-org.libproxy.wlu.ca/topics-policy/projects/online-information-environment>
- Calvillo, D.P., Ross, B.J., Garcia, R.J., Smelter, T.J., & Rutchick, A.M. (2020). Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science, 11*(8), 1119–1128. <https://doi.org/10.1177/1948550620940539>
- Chong, D., & Druckman, J.N. (2007). Framing theory. *Annual Review of Political Science, 10*(1), 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>
- Clark, C.J., Jussim, L., Frey, K., Stevens, S.T., Al-Gharbi, M., Aquino, K., Bailey, J.M., Barbaro, N., Baumeister, R.F., Bleske-Rechek, A., Buss, D., Ceci, S., Del Giudice, M., Ditto, P.H., Forgas, J.P., Geary, D.C., Geher, G., Haider, S., Honeycutt, N., Joshi, H., . . . von Hippel, W. (2023). Prosocial motives underlie scientific censorship by scientists: A perspective and research agenda. *Proceedings of the National Academy of Sciences, 120*(48):e2301642120. doi: 10.1073/pnas.2301642120.
- Clark, J. (2023, March 17). *Twitter files: Vast censor project unleashed against politically incorrect covid-19 speech*. The Washington Times. <https://www.washingtontimes.com/news/2023/mar/17/twitter-files-vast-censor-project-unleashed-against/>
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Conway, L.G., Woodard, S.R., Zubrod, A., & Chan, L. (2021). Why are Conservatives less concerned about the coronavirus (COVID-19) than Liberals? comparing political, experiential, and partisan messaging explanations. *Personality and Individual Differences, 183*, 111124. <https://doi.org/10.1016/j.paid.2021.111124>
- Deppe, K.D., Gonzalez, F.J., Neiman, J.L., Jacobs, C., Pahlke, J., Smith, K.B., & Hibbing, J.R. (2015). Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment and Decision Making, 10*, 314–331.
- Dharshing, S., Hille, S.L., & Wüstenhagen, R. (2017). The influence of political orientation on the strength and temporal persistence of policy framing effects. *Ecological Economics, 142*, 295–305. <https://doi.org/10.1016/j.ecolecon.2017.05.014>
- Ditto, P.H., Liu, B.S., Clark, C.J., Wojcik, S.P., Chen, E.E., Grady, R.H., Celniker, J.B., & Zinger, J.F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science, 14*, 273–291.
- Ditto, P.H., & Lopez, D.F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568–584. <https://doi.org/10.1037/0022-3514.63.4.568>
- Douglas, K.M. (2021). Covid-19 conspiracy theories. *Group Processes & Intergroup Relations, 24*(2), 270–275. <https://doi.org/10.1177/1368430220982068>
- Douglas, K.M., Uscinski, J.E., Sutton, R.M., Cichocka, A., Nefes, T., Ang, C.S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology, 40*(S1), 3–35. <https://doi.org/10.1111/pops.12568>
- Druckman, J.N. (2001). On the limits of framing effects: Who can frame? *The Journal of Politics, 63*(4), 1041–1066. <https://doi.org/10.1111/0022-3816.00100>
- Ecker, U.K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L.K., Brashier, N., Kendeou, P., Vraga, E.K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Enders, A., Farhart, C., Miller, J., Uscinski, J., Saunders, K., & Drochon, H. (2023). Are Republicans and conservatives more likely to believe conspiracy theories? *Political Behavior*. <https://doi.org/10.1007/s11109-022-09812-3>
- Hamilton, L. C., & Safford, T.G. (2021). Elite cues and the rapid decline in trust in science agencies on COVID-19. *Sociological Perspectives, 64*(5), 988–1011. <https://doi.org/10.1177/07311214211022391>

- Hegland, A., Zhang, A.L., Zichettella, B., & Pasek, J. (2022). A partisan pandemic: How covid-19 was primed for polarization. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 55–72. <https://doi.org/10.1177/00027162221083686>
- Imhoff, R., & Lamberty, P. (2017). Too special to be duped: Need for uniqueness motivates conspiracy beliefs. *European Journal of Social Psychology*, 6(47), 724–734. <https://doi.org/10.1002/ejsp.2265>
- Iyengar, S., & Westwood, S.J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690–707. <http://www.jstor.org/stable/24583091>
- Jaidka, K., Mukerjee, S., & Lelkes, Y. (2023). Silenced on social media: The gatekeeping functions of shadowbans in the American twitterverse. *Journal of Communication*, 73(2), 163–178. <https://doi.org/10.1093/joc/jqac050>
- Jansen, S., & Martin, B. (2015). The Streisand effect and censorship backfire. *International Journal of Communication*, 5, 656–671.
- Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–30. <https://doi.org/10.1145/3479525>
- Jolley, D., Meleady, R., & Douglas, K.M. (2019). Exposure to intergroup conspiracy theories promotes prejudice which spreads across groups. *British Journal of Psychology*, 111(1), 17–35. <https://doi.org/10.1111/bjop.12385>
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8, 407–424.
- Kahan, D.M., Peters, E., Dawson, E.C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86. <https://doi.org/10.1017/bpp.2016.2>
- Kubin, E., von Sikorski, C., & Gray, K. (2024). Political censorship feels acceptable when ideas seem harmful and false. *Political Psychology*. <https://doi.org/10.1111/pops.13011>
- McLamore, Q., Syropoulos, S., Leidner, B., Hirschberger, G., Young, K., Zein, R.A., Baumert, A., Bilewicz, M., Bilgen, A., van Bezouw, M.J., Chatard, A., Chekroun, P., Chinchilla, J., Choi, H.-S., Euh, H., Gomez, A., Kardos, P., Khoo, Y.H., Li, M., . . . Burrows, B. (2022). Trust in scientific information mediates associations between conservatism and coronavirus responses in the U.S., but few other nations. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-07508-6>
- Nelson, T.E., Oxley, Z.M., & Clawson, R.A. (1997). Toward a psychology of framing effects. *Political Behavior*, 19, 221–246.
- Nicholas, G. (2022). *Shedding Light on Shadowbanning*. <https://doi.org/10.31219/osf.io/xcz2t>
- Olshansky, A., & Landrum, A.R. (2020). Third-person perceptions and calls for censorship of flat earth videos on YouTube. *Media and Communication*, 8(2), 387–400. <https://doi.org/10.17645/mac.v8i2.2853>
- Palan, S., & Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Pennycook, G., & Rand, D.G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Pereira, A., Harris, E., & Van Bavel, J.J. (2023). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, 26(1), 24–47. <https://doi.org/10.1177/13684302211030004>
- Petrocelli, J.V. (2022). Politically oriented bullshit detection: Attitudinally conditional bullshit receptivity and bullshit sensitivity. *Group Processes and Intergroup Relations*, 25, 1635–1652. <https://doi.org/10.1177/1368430220987602>
- Roberts, M.E. (2020). Resilience to online censorship. *Annual Review of Political Science*, 23(1), 401–419. <https://doi.org/10.1146/annurev-polisci-050718-032837>
- Rogers, R. (2020). Deplatforming: Following extreme internet celebrities to telegram and alternative social media. *European Journal of Communication*, 35(3), 213–229. <https://doi.org/10.1177/0267323120922066>
- Rosenberg, B.D., & Siegel, J.T. (2018). A 50-year review of psychological reactance theory: Do not read this article. *Motivation Science*, 4(4), 281–300. <https://doi.org/10.1037/mot0000091>
- Settle, J.E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.
- Skitka, L.J., & Washburn, A.N. (2016). Are conservatives from Mars and liberals from Venus? Maybe not so much. In P. Valdesolo & J. Graham (Eds.), *Social psychology of political polarization* (pp. 78–101). New York, NY: Routledge/Taylor & Francis Group.
- Sule, S., DaCosta, M.C., DeCou, E., Gilson, C., Wallace, K., & Goff, S.L. (2023). Communication of COVID-19 misinformation on social media by physicians in the US. *JAMA Network Open*, 6(8):e2328928. doi:10.1001/jamanetworkopen.2023.28928
- Taber, C.S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Tucci, L. (2023). Definition: Information Age. *TechTarget*. Retrieved August 2023 from <https://www.techtarget.com/searchcio/definition/Information-Age>
- Twitter files. (2023, August 6). In *Wikipedia*. https://en.wikipedia.org/wiki/Twitter_Files
- Vogels, E.A., Perrin, A., & Anderson, M. (2020, August 19). *Most Americans think social media sites censor political viewpoints*. Pew Research Center. <https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/>
- Vogels, E.A., Anderson, M., Porteus, M., Baronavski, C., Atske, S., McClain, C., Auxier, B., Perrin, A., & Ramshankar, M. (2021, May 19). *Americans and “cancel culture”: Where some see calls for accountability, others see censorship, punishment*. Pew Research Center. <https://www.pewresearch.org/internet/2021/05/19/americans-and-cancel-culture-where-some-see-calls-for-accountability-others-see-censorship-punishment/>
- Washburn, A.N., & Skitka, L.J. (2017). Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science. *Social Psychological and Personality Science*, 9, 972–980.
- Weiss, B. (2018, May 8). *Meet the renegades of the Intellectual Dark Web*. New York Times. <https://www.nytimes.com/2018/05/08/opinion/intellectual-dark-web.html>

- Wilson, A.E., Parker, V.A., & Feinberg, M. (2020). Polarization in the contemporary political and media landscape. *Current Opinion in Behavioral Sciences*, 34, 223–228. <https://doi.org/10.1016/j.cobeha.2020.07.005>
- Worchel, S. (1992). Beyond a commodity theory analysis of censorship: When abundance and personalism enhance scarcity effects. *Basic and Applied Social Psychology*, 13(1), 79–92. https://doi.org/10.1207/s15324834basps1301_7
- Worchel, S., & Arnold, S.E. (1973). The effects of censorship and attractiveness of the censor on Attitude Change. *Journal of Experimental Social Psychology*, 9(4), 365–377. [https://doi.org/10.1016/0022-1031\(73\)90072-3](https://doi.org/10.1016/0022-1031(73)90072-3)
- Worchel, S., Arnold, S.E., & Baker, M. (1975). The effects of censorship on attitude change: The influence of censor and communication characteristics. *Journal of Applied Social Psychology*, 5(3), 227–239. <https://doi.org/10.1111/j.1559-1816.1975.tb00678.x>
- Zarocostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676. [https://doi.org/10.1016/s0140-6736\(20\)30461-x](https://doi.org/10.1016/s0140-6736(20)30461-x)
- Zhao, E., Wu, Q., Crimmins, E.M., & Ailshire, J.A. (2020). Media trust and infection mitigating behaviours during the COVID-19 pandemic in the USA. *BMJ Global Health*, 5(10). <https://doi.org/10.1136/bmjgh-2020-003323>