



Alabama Law Scholarly Commons

Articles

Faculty Scholarship

2023

The Credibility Effect: Defamation Law and Audiences

Yonathan A. Arbel

University of Alabama - School of Law, yarbel@law.ua.edu

Follow this and additional works at: https://scholarship.law.ua.edu/fac_articles



Part of the [Law Commons](#)

Recommended Citation

Yonathan A. Arbel, *The Credibility Effect: Defamation Law and Audiences*, 52 J. LEGAL Stud. 417 (2023).

This Article is brought to you for free and open access by the Faculty Scholarship at Alabama Law Scholarly Commons. It has been accepted for inclusion in Articles by an authorized administrator of Alabama Law Scholarly Commons. For more information, please contact cduncan@law.ua.edu.

The Credibility Effect: Defamation Law and Audiences

Yonathan A. Arbel

ABSTRACT

What should be the legal response to false statements? In the context of defamation law, courts try to set a standard that balances the interests of speakers and their potential targets. This article empirically demonstrates an unappreciated effect of such decisions on third parties: a credibility effect. Using a series of lab experiments, I find that defamation law makes individuals more trusting of reports from various media. This credibility effect is desirable when the report is true but can lead to unintended consequences in the case of misinformation. In particular, the credibility effect is shown to cast a stigma on innocent targets who choose not to file lawsuits. The existence of the credibility effect calls for different balances than are currently employed in defamation law; challenges the vindication justification; and, more broadly, illustrates the limits of policies intended to fight misinformation.

1. INTRODUCTION

How should society regulate false speech? The answer to this question implicates a complex set of judgments of fact and value. In the context of defamation law, courts have approached this question by balancing the interests of two competing parties: speakers and the targets of their speech. Courts seek to set a legal standard that will offer sufficient protection to targets of defamatory speech but will not unduly discourage speakers from participating in public discourse in the marketplace of ideas. What courts neglect to consider in this bilateral model is the possibility that the legal rule affects third parties: audiences. This is a curious omission given that audiences constitute a fundamental part of the mar-

YONATHAN A. ARBEL is Associate Professor at University of Alabama School of Law. I thank Hillel Bavli, Kevin Cope, Shahar Dillbary, Michael Gilbert, Ben McMichael, Cherie Metcalf, Jonathan Nash, and Abraham Wickelgren for helpful comments and advice. I also appreciate the substantive contributions of the two referees and the *Journal's* editors to the development of this paper.

[*Journal of Legal Studies*, vol. 52 (June 2023)]

© 2023 by The University of Chicago. All rights reserved. 0047-2530/2023/5202-0012\$10.00

ketplace of ideas (Heymann 2011). This article finds empirical evidence, consistent with the theoretical predictions in Hemel and Porat (2019) and Arbel and Mungan (2019, 2023), that the law has a first-order effect on audiences. In essence, defamation law produces a credibility effect, which endows speakers and the press with greater public trust. The credibility effect is perhaps desirable when a report is true but leads to perverse outcomes when it is false. As a result, defamation law plays a much broader and nuanced role in democracies than commonly assumed.

Defamation law is a branch of tort law that imposes liability on speakers who publish false allegations that jeopardize a target's standing in the community. Under common law, judges have significant discretion in setting the legal standard by adjusting the level and burden of proof, defining what counts as harmful speech, setting the scope of damages, and recognizing defenses and privileges.¹ In tailoring the legal standard, the jurisprudence of the Supreme Court dictates the bilateral balancing model, seeking to protect victims but not chill broad participation in public debates.²

Recent work has sought to challenge the dominance of the bilateral model. Hemel and Porat (2019), contemporaneously with Arbel and Mungan (2019, 2023), develop theoretical models in which defamation and public trust are linked. Audiences reason, in these models, that if defamation law is effective, speakers will be less inclined to lie for fear of legal sanctions, and so it is rational to place greater trust in their statements. Conversely, weak defamation laws should induce mistrust, which leads audiences to assign lower credence to publications and to rely instead on alternative credibility cues.

If these theories are correct and defamation law carries a credibility effect, then there are several important implications for the design of legal policy. Policy makers interested in saving the sinking public trust in the media could enlist defamation law to foster greater trust. Alternatively, if policy makers seek to diminish the perceived credibility of problematic sources, say tabloids or social media, deregulation could be helpful. More

1. Damage awards in defamation cases can be substantial and involve a great degree of discretion, as illustrated by *Lesher v. Doescher* (2013 Tex. App. LEXIS 12655), in which the court affirmed a \$13.8 million award.

2. As the Supreme Court explains in *Curtis Publishing v. Butts* (388 U.S. 130, 153 [1975]), the law attempts to “strike a fair balance between the interests of the community in free circulation of information and those of individuals in seeking recompense for harm done by the circulation of defamatory falsehood.”

generally, the neglect of credibility effects means that modern standards are likely improperly balanced (Arbel and Mungan 2023).

Why have courts failed to consider audiences in their rulings? There are three main objections to the realism of the credibility effect. The first is informational. Lay audiences are not likely to master the intricacies of defamation law. The second relates to legal consciousness. Even if audiences are generally aware of defamation law, they might not spontaneously consider it when they read the news. These are real concerns, but they are partially mitigated in practice. People frequently consume information through intermediaries—trusted journalists, pundits, or public intellectuals—and these sources are often legally sophisticated. The credibility effect can filter through these intermediaries even if the public is largely unaware of the legal standard.³

The third, and perhaps most challenging, objection is cognitive. The credibility effect is cognitively demanding, because it requires audiences to reason from the perceived incentives of the speaker. A variety of studies, however, show that individuals deviate from the prescriptions of Bayesian decision theory; medical professionals' base rate of neglect in interpreting test results is a prominent example (Casscells, Schoenberger, and Graboy 1978). The inferential challenge from a speaker's incentives is especially difficult given the fundamental attribution problem. According to various studies, individuals systematically ignore the effects of external, situational incentives of others and instead attribute others' actions to their internal dispositions or characters (Ross 1977).

Given the importance to policy of the credibility effect, empirical evidence is needed. This article seeks to offer the first step in this direction. Section 2 provides a brief legal background and a review of the existing, limited evidence for this issue.

Section 3 presents the primary study, a randomized vignette trial of the credibility effect with 200 participants. Respondents were presented with realistic factual statements made in four media—TV news, a newspaper, a blog, and a social media platform—and were asked to evaluate the credibility of the statements. The treatment consisted of the defamation law regime, with half of respondents assigned to an effective regime (where it is easy to prove defamation and recover damages) and half assigned to an ineffective regime. The findings establish the existence of a clear credibility effect: in lab settings the legal regime affects the level of

3. I am thankful to a referee for this insight.

trust that lay individuals place in information provided by the media. Notwithstanding the concerns with cognitive limitations, assignment to an effective defamation regime results in a considerable increase in the perceived credibility of statements. Further bolstering this conclusion, a qualitative analysis of responses suggests that the effect is mediated by a source's higher perceived credibility, as predicted by theory.

Section 4 complements the analysis by considering the vindication effect. A standard justification of defamation law is that it creates a venue for individuals to defend their reputations against malicious attacks. To examine this justification in light of the credibility effect, a reverse vindication question was presented. Respondents were asked to evaluate the credibility of a statement given that its target did not file a lawsuit. Consistent with the existence of a credibility effect, the study finds that subjects in the lab attached a strong stigma to individuals who did not sue. They were more than three times more likely to attach a stigma under an effective defamation regime than under a weak one (a 36-percentage-point difference). The powerful stigma resulting from the credibility effect suggests the subtle nature of balancing in defamation law and an inherent issue with vindication arguments.

The preceding experiments demonstrate the limits of the cognitive objection to the credibility effect. The experiment described in Section 5 seeks to provide preliminary evidence for the informational objections and challenges based on legal consciousness. In that experiment, UK and US respondents ($N = 200$) were asked to rate the credibility of four news reports. After making their assessments, they were asked to evaluate the actionability of the stories, that is, the legal exposure of the publisher if the stories were false. The design allows for inter- and intrasubject analyses, testing for differences in perceptions of actionability between UK and US respondents and the correlation between individuals' assessments of actionability and credibility.

This article has two key findings. First is a surprising convergence in assessments of actionability between respondents from the United States and the United Kingdom. This was unexpected given that UK libel law is considerably more stringent than US libel law. The article considers a few explanations, most centrally that individuals have pervasive and deep misperceptions of the legal standard. It is possible that US participants have an exaggerated sense of defamation law's effectiveness, while UK participants may understate its relevance. The second finding is a strong positive association between judgments about actionability and credibil-

ity spontaneously made by lay individuals. The strength of this finding suggests the relevance of legal perceptions to credibility assessments—subject to the inferential limitations of correlation analysis.

If judgments about credibility are linked to biased perceptions of actionability, individuals likely possess a biased view of the credibility of stories they read in everyday life. The direction and magnitude of misperceptions can amplify or diminish the credibility effect. These misperceptions similarly affect the other stipulated behavioral effects common to the traditional analysis, such as the chilling of speech and vindication effects. As a result, the study of legal consciousness emerges as an important avenue for future research.

Section 6 concludes by discussing limitations and policy implications. The primary policy implication is the need for courts and scholars to design defamation law in light of its broader social effects. If lack of trust is a concern, policy makers could enhance trust by making defamation law stricter by, for example, increasing sanctions, limiting privileges, or reducing the level of proof. At the same time, there are reasons to worry about credibility effects leading to too much trust, as that might engender systemic fragility that can be exploited by adversarial parties. Additional reason for moderation is the stigma effect of defamation law. The existence of common misperceptions suggests that the highest policy payoffs can come not from legal reform but from changes in public communications. Importantly, awareness campaigns do not run into the constitutional and legal hurdles that legal reform invites.

Beyond defamation law, the existence of a credibility effect can inform the study of other domains where the law regulates the veracity of public disclosures, such as false advertising law, securities law, food and drug labeling, Internet providers' liability (section 230 of the Communication Decency Act), and stolen valor. Striking the right balance in defamation law or any of these areas can be difficult. But it will be impossible to do without accounting for the law's broader social effects.

2. BACKGROUND AND MOTIVATION

Defamation law regulates the dissemination of false statements that threaten a target's standing in the community. Under defamation law, a target of a public statement that is false and harmful to his or her good name can sue for all resulting damages. While the doctrine is convoluted

and abstruse, it is held to be “virtually axiomatic” that the law’s overarching goal is the protection of the target’s reputation (McNamara 2008, p. 1).

Pushing against the protection of reputation is the concern that its protection has undesirable effects on speakers. Lay individuals may be reluctant to participate in public debate if defamation law is rigorously enforced. Thus, courts have determined that the legal standard should create a “breathing space” for speakers (*Hustler Magazine Inc. v. Falwell*, 485 U.S. 46, 56 [1988]). The concern with speakers encompasses publishers, and economically oriented scholarship has emphasized how defamation law affects investigative journalism and the media (Garoupa 1999a, 1999b; Bar-Gill and Hamdani 2003; Dalvi and Refalo 2008).

Responding to these competing concerns, the Supreme Court in *Curtis Publishing v. Butts* (388 U.S. 130 [1967]) summarizes the court’s role in defamation cases in terms of a bilateral balancing model. The Court indicates that defamation lawsuits that involve public issues require “accommodation” between the “competing considerations” of open discussion of public matters and “society’s pervasive and strong interest in preventing and redressing attacks upon reputation” (388 U.S. 147, 129).

Until recently, little attention was given to the role of audiences in defamation law. This omission is significant and curious. It is significant because defamatory speech is harmful only if it is believed by audiences, and thus it seems critical to account for the direction and magnitude of any change in audiences’ perceptions of the target. And it is curious because the marketplace of ideas, a key metaphor in First Amendment jurisprudence, is constituted mostly of audiences. Omitting audiences from the analysis creates a deep gap in knowledge about the marketplace. Recent work argues that accounting for audiences requires recognizing potential credibility effects, where stringent defamation standards would increase audiences’ trust (Hemel and Porat 2019; Arbel and Mungan 2019). If true, the credibility hypothesis has a broad range of welfare and policy implications (see Arbel and Mungan 2023).

As noted, there are three primary challenges to the realism of the credibility effect: informational objections, objections based on legal consciousness, and cognitive objections. Audiences may be ignorant of the notoriously complex body of doctrine, may not be aware of the law in everyday life, and even if informed and aware may be cognitively limited in incorporating it into their judgments. There is arguably a difference

between understanding that the law may sanction the speaker for false statements and being able to adjust beliefs on this basis.

While the existence of the credibility effect of defamation law has not been tested, some findings from empirical studies in adjacent contexts are inconsistent with it. Several studies show that audiences do not alter their assessments of credibility when they learn that an advisor is advocating a self-serving position (for example, Cain and Banker 2020). While the majority of studies suggest otherwise, some studies find that low-credibility sources prove as effective as higher-credibility sources in persuading audiences (Pornpitakpan 2004). Studies in the context of food labeling find that consumers exhibit similar levels of trust in disease prevention claims (which are regulated by the Food and Drug Administration [FDA]) and structure claims (which are not regulated), which suggests that the FDA's information regulation does not affect trust in statements (France and Bone 2005).

Empirical work on defamation is limited. Cohen et al. (1988) find in a set of experiments that individuals discount biased defamatory publications but believe that other people will be influenced by them (the third-person effect). Work in communication theory finds that the level of directness in sentences affects the perception of the defamatory nature of statements (Lee 2012).

Given the absence of data and the conflicting theories, it becomes important to test the validity of the credibility hypothesis. Do lay folks consider the legal regime when making credibility judgments?

3. RANDOMIZED CONTROLLED TRIAL OF THE CREDIBILITY EFFECT

3.1. Research Design and Methodology

To examine the effect of defamation law on lay individuals' reasoning, 200 participants were recruited through Positly, an online platform built around Amazon's Mechanical Turk (MTurk). Positly's advantage is that it offers a number of measures of quality control to ensure that respondents are attentive, fit the recruiting criteria, have good standing on the platform, and are indeed unique individuals. The choice of MTurk as a recruiting tool is well analyzed, and—given its quality controls—it is becoming a standard recruiting tool in many areas of research.

The demographics of the sample, relative to the general US population (in parentheses), are 44 percent (50.8 percent) female, median age 35

(38), 77.5 percent (60.4 percent) white, median household income in the \$40,000–\$59,000 range (\$57,652), and 71 percent (30.9 percent) with a college degree or higher level of education. The sample skews white and is significantly more educated than average. There are no theoretical predictions as to the effect of race. Education levels may be associated with a greater propensity to engage in higher-level abstraction and calculations. This could lead to the sample overweighting the relevance of defamation law relative to the general population. Still, post hoc analysis does not identify any significant (economically or statistically) correlations.

The treatment consisted of assigning respondents to different defamation law regimes. The challenge was to make the governing law clear to respondents, and this could not be done without both informing them about the relevant law and priming them to its existence. The inferential limitations imposed by this structure are discussed below, but the primary goal of this study should be understood as testing the existence of the credibility effect in light of contradictory predictions instead of offering a precise measurement of its magnitude.⁴

Respondents were given a short explanation about the governing law in the imaginary state where the events take place (reproduced in Section OA1 of the Online Appendix). In the group assigned to the effective defamation law regime, respondents were told that people who make false accusations, and only them, are likely to be sued and made to pay substantial damages to their victims. In the group assigned to the ineffective law regime, a similar statement was made in the negative, that is, people who make false allegations are unlikely to be sued and, if so, unlikely to pay substantial damages. This treatment sets a large difference in legal regime between the groups, which was considered suitable to detect the existence of an effect.

The standard bilateral model employed by the court neglects audiences and thus implicitly assumes that defamation law does not have important third-party effects. On this account, one would expect there to be little effect of the group assignment on assessments of credibility. This is particularly relevant if individuals are incapable—for a variety of cognitive limitations—to factor in the speakers' incentives and update on that basis. A competing hypothesis, provided by the theoretical work noted

4. The interest in assuring internal validity is another reason the design does not include an untreated group. Note, however, that inclusion of such a group would not allow me to cleanly compare control and treatment, as the treatment involves more than just a change in the legal regime.

above, is that the level of defamation law should have a positive effect on judgments about credibility. Thus, the main hypothesis tested here is whether defamation law creates a credibility effect.

After being presented with their assignments, participants were asked a series of attention check questions (93 percent of participants passed, and those who did not were screened out). Participants were then presented with four scenarios, each on a new page so that they could not scroll back.

The scenarios were made realistic to alleviate external validity concerns. Two scenarios involved an allegation by a man against a woman, the third was a man against a man, and the fourth was a firm against a woman. The names were fictitious, and most of the images were computer generated and not of real individuals. Each scenario used a different information medium, and each involved allegations sufficiently serious to make defamation law potentially relevant.

3.1.1. Newspaper Article. Although on the decline, traditional news reporting is still a major source of information. Given the centrality of newspapers in reporting the news, it is not a coincidence that they stand at the center of many defamation law cases.⁵ Thus, the first task presented respondents with a scenario involving a scan of a (researcher-designed) news article (see Figure 1), which details the financial problems of a local firm. The story mentions, without supporting evidence, that some financial problems are due to embezzlement by the company's accountant, Ms. Miranda Dewey.

After an attention check, respondents were presented with a credibility rating question, which asked them to rate on a sliding scale of 0 to 100 the probability that Miranda Dewey embezzled money. Respondents were also asked to explain, using free text, what led them to this decision. They were then asked how likely Dewey would be to win a lawsuit if the allegation is false.

3.1.2. Social Media Post. Individuals consume a great deal of information through social media. While there is sometimes personal familiarity with the individuals involved in a post, it is not uncommon to encounter posts by unfamiliar individuals. This raises several novel challenges, given the low barriers to entry for the reporting of facts on these platforms. Accordingly, the next scenario involved a post on one social media plat-

5. Of the 45 Supreme Court defamation cases, 22 involved newspapers, magazines, or newsletters; this is the largest subject category.



Figure 1. Newspaper article



Figure 2. Social media post

form: Facebook (see Figure 2). The scenario detailed an exchange of allegations between an ex-employee and her former place of employment. While both parties make allegations in this scenario, the focus in the analysis is on the employer's claims. The reason is that the employee may be perceived to be judgment proof, which would render defamation law irrelevant.

Respondents were asked to rate, on a 0–100 scale, the likelihood that Patricia Sutton stole from her employer. Again, the opportunity to ex-



Figure 3. Television news report

plain was given, alongside a rating of how likely she would be to win a lawsuit if the allegation is false.

3.1.3. Television News. Television reporting is a markedly different industry from print news, and it raises some special considerations. Television news involves a dynamic visual experience that is lacking in the consumption of print media. The next scenario involved an artificial image designed to appear to be taken from a TV news report (see Figure 3). The scenario is meant to evoke general perceptions of TV reporting but uses a neutral name for the station (NP Local News) to avoid the polarizing valence attached to channels such as Fox News or MSNBC.

Subjects were told that they were watching a local news station when they came to this segment. They were told that Paul Stevens is a local businessman and that Jim Lane—the defamed—is a local contractor. The setup is meant to elicit some doubt about ulterior motives that might lead the speaker to lie. As before, a 0–100 scale was used to assess the credibility of the report.

3.1.4. Blog Post. Individual media channels, such as blogs, straddle the line between social media and news channels. There is a great degree of heterogeneity among blogs, so one cannot hope to make effective generalizations, but the domain appears sufficiently interesting to merit an investigation. Thus, the final scenario involved a blog post accusing Linda Johnson of attempting to collect compromising information about the local mayor to get leverage over him in an implied blackmail scenario (see Figure 4). The same credibility assessment was repeated.



Brandon Trovney • 2 minutes ago • 1 min read



More New Port Revelations

Just when it seemed like New-Port politics couldn't become any dirtier, my investigations uncovered new and very worrisome details about our local School Board Official, Robert Troy. It turns out that his personal assistant, Linda Johnson, assisted him with some shady dealings. Linda planned to hire a PI to spy on the Mayor. Even worse, Linda planned to get the mayor in some compromising positions, so she could get leverage on him.

Linda is a highly manipulative, corrupt individual. I know her because she was my neighbor for six long years, and she was nothing but trouble. We got into a lot of arguments and I got to see what kind of person she is first hand.

... Read More



Figure 4. Blog post

3.2. Findings

Figure 5 shows, for each scenario, the mean credibility assessment in the effective and the ineffective defamation law regimes. Comparison of the means shows that, in each scenario, the mean credibility assessment of the report is considerably higher in the effective defamation regime, with an increase of between 14 percent and 26 percent. The differences between the groups are all statistically and practically significant.⁶

On average the mean credibility assessment was 54.6 in the ineffective-law treatment and 65 in the effective-law treatment, for a difference of 10.4 percentage points (an increase of 19 percent). On the scales used, 50 denotes a judgment that the allegation is equally likely to be true or false, whereas 75 denotes that it is probably true, which makes the difference in means large and practically significant. The effect is also statistically significant.⁷ However, the size of the effect should be interpreted in the context of a stark change in legal regimes.

Using controls for a variety of demographic factors, Table OA1 in the Online Appendix reports the effects of defamation law on assessments

6. The difference between the groups for the newspaper article is $t(184) = 3.42, p < .01$; the difference for the social media post is $t(184) = 2.27, p < .05$; the difference for the TV news report is $t(184) = 4.75, p < .01$; and the difference for the blog post is $t(184) = 2.95, p < .01$. Including inattentive respondents as a robustness check has little effect on these results. In Figure 5 the means become 72.7 (-.8), 63.7 (1) for the newspaper article; 58.9 (-.3), 51.6 (-.4) for the social media post; 64.9 (-1.2), 52.3 (-.2) for the TV news report; and 60.8 (-.3), 51.9 (.7) for the blog post.

7. The difference is $t(742) = -6.3, p < .001$.

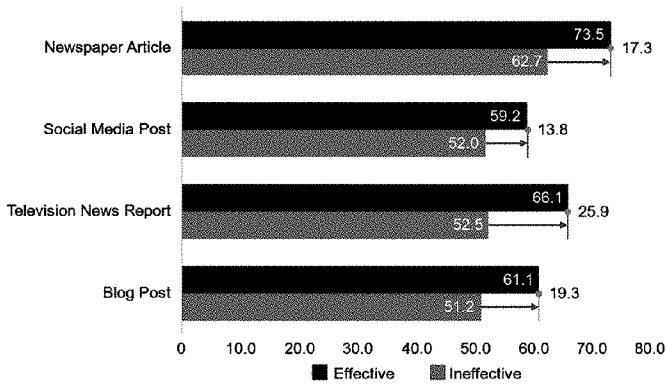


Figure 5. Mean credibility differences by defamation law regime

of credibility across all conditions. Because education can be a source of potential bias, the analysis controls for respondents' education and the interaction of education and treatment group alongside other relevant demographic factors such as age, race, gender, and income. None of the covariates have a statistically significant effect. Even after including these controls, the effect of defamation law on credibility remains large and highly statistically significant.

3.3. Qualitative Analysis

I learned about the respondents' mental models by asking them to explain their assessments of credibility using free text. Overall, across the scenarios, the general impression is of a bifurcated inferential role for defamation law. In the effective-law group, respondents rely on defamation law as a credibility cue, using it to ground their decision whether to believe the communication. In the ineffective-law group, however, respondents rely more heavily on alternative credibility cues, and defamation law is mentioned more in passing.

To illustrate, in the context of the newspaper article, members of the effective-law group make the following assertions (the corresponding credibility assessment is in parentheses): "I don't think a paper would take the risk without researching into a story first, especially something this big. If they did fail to post the truth, they would lose a lot of money in a defamation lawsuit because Miranda's name will be forever tarnished" (70); "The New-port news paper knows of the penalty for publishing false information, therefore they would certainly be careful and

make sure the information is true before they publish it. If they had not done their own investigating before they published the information then they will surely get sued and lose" (95); "I'd like to think that the people there know that defamation is likely to get them in trouble so calling out a person for a crime or something like that they better be sure the person did it" (75).

By contrast, the responses in the ineffective-law group are more agnostic and seek other credibility cues, either in the report or in norms of journalistic integrity. One respondent writes, "The reason the company is doing poorly is due to lack of resources not because of embezzlement" (20); another states, "I picked 75% likely because if a newspaper reports it, they usually try to get the facts right, but sometimes are wrong" (75). Still, a small number of respondents draw a negative inference from defamation law, noting that "because defamation laws are so lax, there is also a large possibility that they got the facts wrong and didn't check because there are no real consequences for reporting false news" (49).

The alternative credibility cues to defamation law took different forms. In the social media scenario, for example, respondents note the unreliability of social media, the relevance of personal familiarity, and the role of reputation. One respondent stated, "It's a facebook post, anything you read on social media is about as accurate as flipping a coin" (50); another reasons that "given that I'm seeing her in my feed, I imagine I would know her, and most likely would believe her denials, if I trusted her enough to allow her to friend me" (30); another respondent noted that "even in this state, I don't think a company would risk the negative PR of lying about an employee" (60). While no respondent made explicit inferences based on race or sex, two arguably came close: "I feel as if the woman is trying to blame the company, when it was her fault" (23); "It sounds like she's a bitter ex-employee" (60).

Finally, participants incorporated rich pragmatic legal considerations as mediating factors. For example, in the social media scenario, members of the effective defamation law group considered legal awareness: "I think a company would know more about the laws and not want to risk a lawsuit. I don't think they would risk saying it if it was not true" (100). Participants went beyond the odds of winning a lawsuit and considered the probability of one being brought: "Patricia doesn't have a lot of money to pay lawyers" (80); "Sutton would not have the funds to file a defamation lawsuit" (80). Similarly, another respondent considers the disutility of suing Patricia: "Even if Patricia Sutton was lying, suing her

for defamation would likely not result in any gain for FRI. So, I would believe they are the ones telling the truth here, and Patricia is lying” (50). This suggests that lay individuals can be legal realists and that their judgments incorporate not only nominal legal norms but also a richer set of pragmatic legal considerations.

3.4. Summary

Taken together, these findings show the causal effect of defamation law on assessments of credibility. Moving from an ineffective-law regime to an effective one leads to a 19 percent change in respondents’ assessments of credibility. This effect was true for all media types (newspaper, social media, TV news, and blog). Overall, people exhibited greater trust in the media when defamation law was made more effective. In the abstract, the effect size is significant, although in some instances the practical effect of such a change can be minimal (if, for example, the statement is deeply distrusted) or pivotal (if individuals are teetering with their trust of a source or a statement).

Analysis of the responses further reinforces the role of defamation law on lay judgments of credibility. It also revealed, somewhat unexpectedly, the differential role that defamation law played in respondents’ analysis. In the effective-law group, respondents relied more heavily on defamation law as an important factor in forming judgments. In the ineffective-law group, however, respondents relied more on alternative cues of credibility and drew fewer explicit inferences from the weakness of defamation law. The substitution of legal cues with nonlegal cues may be consequential from a social perspective, as it was hinted by some respondents that weak defamation laws can invite greater reliance on speakers’ personal demographics such as sex, although the evidence on this score is very preliminary. It further appears that individuals are realists in the sense of caring not only what the law says but also how likely a lawsuit is to be brought and how aware the parties are of the law. Against the animating concern that individuals lack the cognitive aptitude or sophistication to make conditional judgments based on the legal regime, this study demonstrates the propensity of lay individuals to adjust their beliefs on the basis of defamation law.

A number of limitations that affect the interpretation of these findings are summarized in Section 5.3. The most important one is that respondents in the study were informed about the content of defamation law and, consequently, may have been primed to think about it.

4. CREDIBILITY EFFECTS AND VINDICATION

The Supreme Court states that “an action for damages is the only hope for vindication or redress the law gives to a man whose reputation has been falsely dishonored” (*Rosenblatt v. Baer*, 383 U.S. 75, 93 [1966] [Stewart, J., concurring]). Two leading tort scholars explain that an important justification for defamation law is that it provides “right to vindication in the face of . . . deliberate and malicious attacks by others” (Goldberg and Zipursky 2020, p. 338).

The theory animating these statements is that defamation law opens a venue for victims who can vindicate their reputations by receiving a supporting judgment. Vindication thus provides an independent justification for defamation law.

The possibility of a credibility effect suggests a competing theory (Arbel and Mungan 2019). The vindication theory expects the public to update its perceptions of a victim when a redeeming judgment comes to light. The same public should also be able to observe the failure of a victim to bring suit (a fact the speaker may strategically advertise). The public may then attach a stigma of guilt to those who fail to bring suit, even though there may be legitimate reasons for a victim not to sue. The stronger defamation laws are, the greater the stigma attached to failing to bring a lawsuit.

To explore this reverse vindication effect, the randomized controlled trial reported above ($n = 186$) included a vindication question. In the blog post scenario, in which Linda Johnson was accused by a local blogger of attempting to blackmail the mayor, respondents were asked to evaluate the claim’s credibility. They were then asked to rate, on a 0–100 scale, the likelihood that the following proposition is true: if Linda Johnson fails to sue the blogger, that means she is guilty of the accusation.

Figure 6 reports the findings of the analysis. The mean perception of the victim’s guilt from failure to bring a lawsuit in the effective-law group is 52.6, relative to 16.2 in the ineffective-law group. The difference is highly statistically and practically significant.⁸ The stigma associated with failure to bring suit is small when defamation laws are weak, but it becomes very strong under effective defamation laws. Defamation law thus casts those who do not sue in a negative light.

8. The difference is $t(184) = 7.98$, $p < .01$.

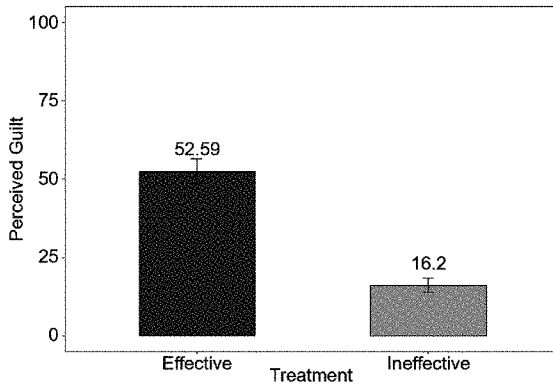


Figure 6. Stigma effect of defamation laws

5. CREDIBILITY EFFECTS IN ORDINARY JUDGMENTS

How large is the credibility effect outside the lab? Theoretical considerations work in opposite directions. In everyday life, individuals lack information about the law and may not be primed to think about it. On the other hand, the credibility effect can also filter indirectly. Individuals are likely not aware of the level of enforcement by the FDA, but if they observe over time that FDA-approved drugs are safe, they will learn to trust drugs with an FDA label. Similarly, sophisticated intermediaries, who are aware of the legal standard, may adjust their publication standards accordingly. By observing the intermediary's track record, the public can adapt its assessments of credibility.

Another channel for the credibility effect is direct: people respond to (their perceptions of) the legal standard. The present study attempts to shed additional light on this direct channel. It does so by first measuring the difference in legal consciousness between lay participants in jurisdictions where the legal standard is markedly different—the United States and the United Kingdom. Second, it measures the implicit association between legal perceptions and judgments about credibility.

5.1. Research Design and Methodology

To test these hypotheses, participants from the United States (101 people) and the United Kingdom (99 people) were recruited through the service Prolific. The demographics of the US sample, relative to the general US population (in parentheses), are 39.6 percent (50.8 percent) female, me-

dian age of 32 (38), 67.3 percent (60.4 percent) white, median household income of \$72,000 (\$57,652), and 76.2 percent (39.1 percent) with a college degree or higher level of education. The most significantly skewed category was education, and the sample is otherwise somewhat more wealthy, male, young, and white than the general population. With the exception of education, none of these factors are theoretically expected to affect the findings. In the UK sample, the demographics are 67.7 percent (50.5 percent) female, median age of 32 (40.5), 97 percent (87.2 percent) white, median household income of \$46,900 (\$40,484), and 59.6 percent (38.6 percent) with a college degree or higher level of education. The most significantly skewed category is again education, although it is skewed to a lesser extent than in the US sample. The UK sample is somewhat younger, whiter, and wealthier than the general population. There is also a gender skew in this sample, but this time females are overrepresented.

The participants were asked a few demographic questions, moved to a page with an attention check question, and then given four scenarios.⁹ One is a newspaper clipping detailing a story about the financial distress of a local firm. The second is a social media post detailing an exchange between an employee and her former employer. The third is an image from a television report, with a chyron quoting and describing accusations. And the fourth is a short block post sharing a report about local politics. Each scenario described an allegation of doubtful veracity, and respondents were asked to rank, on a 0–100 scale, the likelihood that the statement was true. Critically, there was no mention of defamation law, or any legal concept, in the presentation.¹⁰ Section OA2 in the Online Appendix includes the scenarios.

The scenarios were designed with a few principles in mind. The situations are familiar and diverse. In all of them, the allegations are factually ambiguous, and judgments have to be made from limited information. Even though defamation law is not mentioned, the scenarios would be handled differently under defamation law. For example, one scenario includes a public figure and another includes a private figure, and the former scenario is far less actionable under US law.

9. All participants passed the attention check question.

10. Per institutional review board standards, the invitation to the survey identified the lead researcher and his affiliation, which is a law school. This was the only reference to law at this stage, and it was part of a long disclosure.

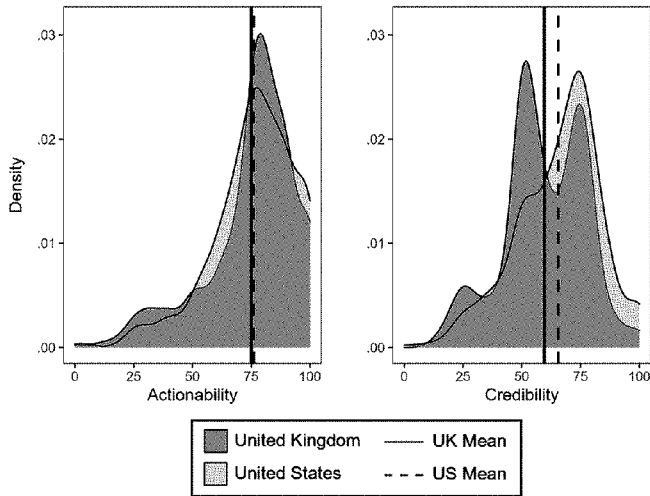


Figure 7. Density plot of judgments about actionability and credibility

After making assessments about credibility, respondents were moved to the next stage and were unable to go back to revise their responses. In the second stage, they were told that the allegation is false and were asked how likely the plaintiff is to prevail in litigation assuming he or she has reasonable evidence. That is, respondents were asked to make an assessment of actionability, which they ranked on a 0–100 scale.

5.2. Findings

The responses were aggregated into 800 response pairs (credibility and actionability). The analysis of the responses yields two key findings, shown in Figure 7. First, surprisingly, UK respondents made judgments about actionability that were almost indistinguishable from those of their US counterparts: the mean assessment was 75.04 among UK respondents and 76.03 among US respondents—a statistically insignificant difference.

One explanation for this convergence is that it is an artifact of the survey design. Respondents may be subject to acquiescence bias in responses (Bentler, Jackson, and Messick 1971) and thus gravitate to an answer they think is expected of them, which might happen to be 75 in this instance. The distribution of assessments of credibility does not support these concerns. Unlike actionability, the difference in mean assessments of credibility between the two groups is large and statistically significant

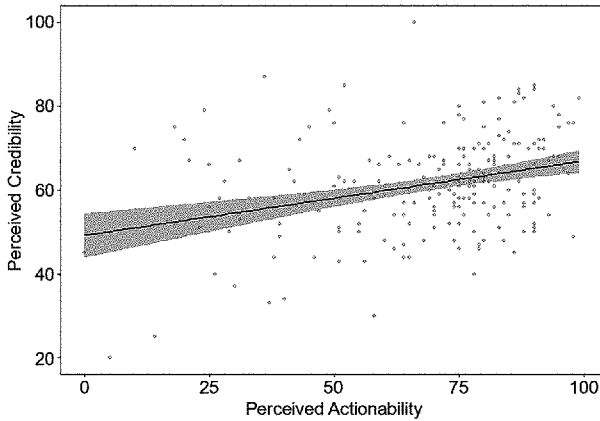


Figure 8. Perceived source credibility and actionability

(65.5 for US respondents versus 59.5 for UK respondents),¹¹ and the distribution of responses is markedly different between the samples.¹²

Figure 8 shows that individuals' assessments of credibility are strongly correlated with their assessments of the legal exposure of the speaker, that is, the statement's actionability. Across all scenarios, there is a correlation of $r = .21$ between the perceived credibility and perceived actionability of a statement. The relationship is highly statistically significant.¹³

The association is also highly statistically significant in three of the four scenarios.¹⁴ While the correlation ranges between .18 and .27 for these three scenarios, it is close to 0 for a newspaper article about a public figure. This last finding is unexpected but may be explained by the relative immunity given in the United States to the media on issues of public concern.

Table OA2 reports the full results of a regression analysis. The association between actionability and credibility holds while controlling for other demographic factors. Beyond the jurisdictional effect, education and gender also play a role. Inasmuch as education is a proxy for so-

11. The difference is $t(797) = 4.8$, $p < .001$.

12. The result of two-sample Kolmogorov-Smirnov tests of the credibility responses for both countries is $D = .194$, $p < .001$.

13. The relationship is $t(798) = 6.14$, $p < .01$.

14. The values for the bulletin board note are $r = .2$, $t(198) = 2.88$, $p < .01$; the values for the company's email message are $r = .27$, $t(198) = 3.93$, $p < .01$; the values for the newspaper article about a public figure are $r = .05$, $t(198) = .75$, $p = .45$; and the values for the newspaper article about a private figure are $r = .18$, $t(198) = .75$, $p = .01$.

p sophistication, a plausible explanation is that sophisticated individuals are better positioned to make more complex assessments of credibility. The effect of gender is unclear and should be a topic of future exploration.

5.3. Discussion

What might explain the convergence of assessments of actionability between two jurisdictions thought to be far apart? The most direct explanation is that misperceptions of defamation law are common. This would not be unexpected: defamation law is notoriously complex and incoherent. Consequently, US respondents may exaggerate its effectiveness, while UK respondents may underestimate its effectiveness, or it may be that individuals converge on general expectations drawn from morality or social norms.

Another possibility is compensating differentials. As the results for the credibility effect suggests, laypeople may be realists and think of legal norms not just nominally but in terms of their overall, practical effect. Thus, the stringent UK standard may be diluted by, for example, a less litigious society, lower damages awards in UK courts, or the English rule on litigation costs. On net, the differences between the systems may be smaller than the literature suggests.

If these misperceptions are real, there are several implications for the study of defamation law. If people in a jurisdiction exaggerate the law's effectiveness, they may become overly trusting or overly skeptical (on the effects of consumers' misperceptions, see Bar-Gill and Davis [2017]). It is important to understand the source of the misperceptions. If people draw their expectations about the law from their prior beliefs or moral judgments, changes in legal policy may have muted behavioral effects on the public. In this case, legal reform would not (directly) affect judgments about credibility, but it might also not chill public speech or affect recourse and vindication.

To be sure, the credibility effect does not depend on direct channels. It was noted that the behavioral effects of defamation law are often mediated by sophisticated intermediaries who are aware of the law and respond to it. Despite the indirect channels, there is still an important reason to study the perceptions of laypeople. Social media has weakened traditional intermediaries; today, lay individuals routinely assume the positions of public speakers, victims, and listeners. Thus, misperceptions can influence judgments about credibility, willingness to engage in public debate, and the ability to bring suit.

The second key result is that respondents spontaneously assigned greater credibility to statements that they thought were more actionable. In other words, respondents believed speakers more when the speakers stood to lose from lying. Importantly, the experiment eliminated references to defamation law when measuring judgments about credibility. Thus, this finding suggests that, of their own accord, individuals consider the legal environment relevant to the evaluation of a speaker's credibility.

To be clear, the study measures only the correlation of perceptions and judgments. In addition, the effect is moderate, which suggests that the effect of defamation law is not a dominant factor. This is expected given the plethora of alternative cues about credibility besides defamation law, such as the plausibility of the claims or the medium. At the same time, there is reason to believe that the true effect is in some sense much larger. If some respondents are overconfident in their assessments of credibility, they may be biased to evaluate the statements as less actionable so as to avoid cognitive dissonance between statements they believe to be true and the possibility that a court will rule that they are not true.

6. CONCLUSION: LIMITATIONS AND POLICY IMPLICATIONS

6.1. Credibility Effects Inside and Outside the Lab

The findings here show that, in lab settings, individuals have both the cognitive capacity and an epistemic motivation to make assessments of credibility based on defamation law. The effect is quite large and robust. Further, individuals attach a strong stigma to plaintiffs who fail to sue. All of this suggests that there is clear potential for credibility effects in everyday judgments.

It is difficult to assess the scope of the credibility effect outside the lab. Part of the problem comes from the finding that there are apparently broad misconceptions about the law. Such misconceptions can lead individuals to experience muted credibility effects or highly exaggerated ones. In the context of consumer contracts, for example, studies find that lay individuals overestimate their enforceability (Wilkinson-Ryan and Hoffman 2015; Furth-Matzkin and Sommers 2020). Scholars worry that such misperceptions lead to overcompliance with contracts that are not enforceable. Similarly, a survey found that 43 percent of consumers erroneously believed that only drugs that were completely safe could be adver-

tised to them (Bell, Kravitz, and Wilkes 1999). This belief, presumably, led consumers to purchase unsafe drugs.

Because misperceptions can magnify behavioral effects in unexpected directions, it becomes important for future work to study the direction, magnitude, and source of lay perceptions of the law. The findings also show that, whatever their misperceptions are, individuals link judgments about credibility and actionability. Thus, it is quite possible that credibility effects lead to heterogeneous and biased judgments in everyday life. At the same time, credibility effects can also flow indirectly through information intermediaries who can be legally sophisticated.

Ultimately, one cannot be sure that individuals who read the news incorporate their perceptions of the law into their unconscious system 1 judgments about credibility (Kahneman 2013). Still, the literature on the credibility of sources (Pornpitakpan 2004) shows that individuals are highly sensitive to subtle and complex cues regarding credibility. The present article finds strong evidence that, when informed, individuals are motivated to rely on defamation law as a cue of credibility and that there is some evidence that they do so of their own accord. In addition, the credibility effect can flow indirectly, through sophisticated intermediaries. For all of these reasons, credibility effects are important to the evaluation and development of the legal standard. The study of lay (mis)perceptions further emerges as an important social question about the law's development.

6.2. Limitations

In addition to the limitations mentioned above, a few others constrain the interpretation of this study. The samples are not representative and skew more educated than the general population, among other differences. In the experiment, education is linked to a greater association between perceptions of credibility and actionability. Thus, the findings might exaggerate the size of the effect. At the same time, the effect of education can be taken to reinforce the underlying hypothesized mechanism. Because consideration of speakers' incentives requires a certain degree of sophistication, education should be expected to play a role.

Another issue involves the scenarios themselves. A routine concern in vignette studies is that the context-rich fact pattern might elicit responses to factors unrelated to the research question. For example, people may perceive statements against women to be both more credible and more

actionable. The use of four scenarios in each study with variation in their contextual factors (such as the speaker's gender) mitigates some of these concerns but cannot eliminate them. The contrary concern is also relevant. The vignettes offer fewer cues about credibility than are available in real-life scenarios, where individuals are able to search for more information. The findings, then, should be interpreted as reflecting scenarios in which limited information exists and individuals are either unable or unmotivated to search for more information.

A final issue concerns the lack of a clearly defined control group. The primary study compares results from two treated groups, one with a lax defamation law and another with a strict one. As a result of the stark difference between the legal regimes, the size of the credibility effect (as distinguished from its existence) should be interpreted with some caution, as legal reform is often incremental and small in scope. Notwithstanding these limitations, the present article provides the first evidence about an important social question on which key evidence is lacking.

6.3. Relevance for Legal Policy

The existence of the credibility effect informs legal policy along several dimensions. The robustness of democracies is often attributed to some notion of a marketplace of ideas and a functioning media, the fourth estate. Mistrust in information provided by others and lack of trust in the press are threats to self-governance. At the same time, trust is a complex concept, built and lost by a variety of factors, many of which are outside the reach of policy makers. The credibility effect suggests a policy lever within policy makers' control: defamation law. Increasing the media's and speakers' accountability through stronger defamation laws could enhance public trust. Of course, this is not the only tool, but given the difficulty and importance of the problem, it is one worth considering.

Credibility effects can flow directly from the law to the public or indirectly through sophisticated intermediaries. When the focus is on direct effects, misperceptions about the law pose a challenge. If individuals are unaware of the legal standard, changing laws will have little effect. But this points to a different policy lever. Policy makers can influence public trust by informing the public of the legal standard. Such campaigns can also prime individuals and make them more likely to consider the legal environment in everyday life. Unlike changes to the legal standard, public awareness campaigns are not subject to First Amendment and other legal constraints.

The credibility effect also suggests the limits on defamation law. Reputational harms come only from people believing derogatory statements about others. If defamation laws are too strict (or are perceived to be such), they can counterintuitively amplify the harm of false statements because of the enhanced trust they create (even if they reduce the overall volume of false speech). The finding that credibility effects also stigmatize those who do not file lawsuits further reinforces the concern with strict defamation laws. These considerations suggest an important trade-off between the credibility of speech and harm from statements. Elsewhere I investigate the welfare implications and provide an argument for moderation and erring on the side of free speech (Arbel and Mungan 2023).

Doctrinally, the credibility effect suggests that the bilateral balancing model employed by courts fails to capture the full potential impact of the legal standard. Courts are balancing two factors when they should balance three. To be sure, the size of the credibility effect may be small if the public is unaware of changes to the legal standard. But so will be the chilling of lay speech and the opportunity for victims to redeem themselves in the eyes of the public. In the future, courts should be more attentive to defamation law's direct and indirect credibility effects. The decision-making process in such cases should factor in the preexisting and desired level of trust in a given domain and the law's public salience. Other things being equal, consideration of credibility effects can tip the scales in one direction in the context of social media and in the other for printed news.

It should be recognized, however, that people may not always be guided by truth-seeking motives when they adopt beliefs. Some beliefs may be held for social rather than epistemic reasons, and so credibility effects may work in unexpected ways. Charlatans often use negative coverage as supposed evidence of prosecution by entrenched interests, which their followers take as reinforcing the charlatan's message. Such dynamics suggest the delicate nature of regulating speech and indicate that development of better empirical understanding of audience dynamics is needed.

On a broader level, in a large number of domains the law seeks to regulate the veracity of information and sanction those who spread falsities. Examples include securities regulation, false advertising, false lights, stolen valor, fraud, trademark, and perjury. Scholars and policy makers across this broad range of diverse contexts have neglected to consider credibility effects and how the regulation of falsities can be too strict.

Future work might report on natural experiments that allow discernment of the causal effect of changes to defamation law on measures of public trust. It would also be productive to learn more about misperceptions that ordinary folks have about the law: do they overestimate the scope of the media's liability? Do they underestimate liability on social media? How salient are changes to defamation law standards? With an accumulation of calls for reform, evidence on these basic questions becomes critical.

REFERENCES

- Arbel, Yonathan A., and Murat Mungan. 2019. The Case against Expanding Defamation Law. *Alabama Law Review* 71:453–97.
- . 2023. Defamation with Bayesian Audiences. *Journal of Legal Studies* 52:445–83.
- Bar-Gill, Oren, and Kevin E. Davis. 2017. (Mis)perceptions of Law in Consumer Markets. *American Law and Economics Review* 19:245–86.
- Bar-Gill, Oren, and Assaf Hamdani. 2003. Optimal Liability for Libel. *Contributions in Economic Analysis and Policy* 2, pp. 1–26.
- Bell, Robert A., Richard L. Kravitz, and Michael S. Wilkes. 1999. Direct-to-Consumer Prescription Drug Advertising and the Public. *Journal of General Internal Medicine* 14:651–57.
- Bentler, P. M., Douglas N. Jackson, and Samuel Messick. 1971. Identification of Content and Style: A Two-Dimensional Interpretation of Acquiescence. *Psychological Bulletin* 76:186–204.
- Cain, Daylian M., and Mohin Banker. 2020. Do Conflict of Interest Disclosures Facilitate Public Trust? *AMA Journal of Ethics* 22:E232–E238.
- Casscells, Ward, Arno Schoenberger, and Thomas B. Graboys. 1978. Interpretation by Physicians of Clinical Laboratory Results. *New England Journal of Medicine* 299:999–1001.
- Cohen, Jeremy, Diana Mutz, Vincent Price, and Albert Gunther. 1988. Perceived Impact of Defamation: An Experiment on Third Person Effects. *Public Opinion Quarterly* 52:161–73.
- Dalvi, Manoj, and James F. Refalo. 2008. An Economic Analysis of Libel Law. *Eastern Economic Journal* 34:74–94.
- France, Karen Russo, and Paula Fitzgerald Bone. 2005. Policy Makers' Paradigms and Evidence from Consumer Interpretations of Dietary Supplement Labels. *Journal of Consumer Affairs* 39:27–51.
- Furth-Matzkin, Meirav, and Roseanna Sommers. 2020. Consumer Psychology and the Problem of Fine-Print Fraud. *Stanford Law Review* 72:503–60.

- Garoupa, Nuno. 1999a. Dishonesty and Libel Law: The Economics of the “Chilling” Effect. *Journal of Institutional and Theoretical Economics* 155:284–300.
- . 1999b. The Economics of Political Dishonesty and Defamation. *International Review of Law and Economics* 19:167–80.
- Goldberg, John C. P., and Benjamin C. Zipursky. 2020. *Recognizing Wrongs*. Cambridge, MA: Belknap Press of Harvard University Press.
- Hemel, Daniel, and Ariel Porat. 2019. Free Speech and Cheap Talk. *Journal of Legal Analysis* 11:46–103.
- Heymann, Laura A. 2011. The Law of Reputation and the Interest of the Audience. *Boston College Law Review* 52:1341–1439.
- Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
- Lee, Sungbom. 2012. A Pragmatic Analysis of Defamation and Slandorous Remarks. *Korean Journal of Linguistics* 37:401–16.
- McNamara, Lawrence. 2008. *Reputation and Defamation*. New York: Oxford University Press.
- Pornpitakpan, Chanthika. 2004. The Persuasiveness of Source Credibility: A Critical Review of Five Decades’ Evidence. *Journal of Applied Social Psychology* 34:243–81.
- Ross, Lee. 1977. The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. Pp. 10:173–220 in *Advances in Experimental Social Psychology*, ed. Leonard Berkowitz. New York: Academic Press.
- Wilkinson-Ryan, Tess, and David A. Hoffman. 2015. The Common Sense of Contract Formation. *Stanford Law Review* 67:1269–1301.

