

The Credibility Effect: Defamation Law and Audiences

Yonathan A. Arbel¹

¹School of Law, University of Alabama

What should be the legal response to false allegations? 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 potential effect of such decisions on third parties: a credibility effect. Using a series of lab experiments (total N=400), it finds that defamation law makes individuals more trusting of reports on different media. This credibility effect is desirable when the report is true, but can lead to unintended consequences in the case of misinformation. The credibility effect also casts an unexpected stigma on innocent targets who choose not to file a lawsuit. The existence of the credibility effect calls for different balances than are currently employed in defamation law and, more broadly, illustrates the limits of policies intended to fight misinformation.

Drafts: comments welcome by email yarbel@ua.law.edu or anonymously through bit.ly/arbeleranon.

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 within the ‘marketplace of ideas.’ What courts neglect to consider is the possibility that the legal rule also affects third parties: the audiences of speech. This is a curious omission given that audiences constitute a fundamental part of the marketplace of ideas (Heymann, 2012). This paper finds empirical evidence that the law has a first-order effect on audiences. In essence, defamation law produces a ‘credibility effect,’ endowing speakers and the press with greater public trust. The credibility effect is perhaps desirable when the report is true, but leads to perverse outcomes when it is false. As such, defamation law proves to play 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 their target’s standing in the com-

Draft, Vol. 0, No.0,

[doi:/ewmxxx](https://doi.org/10.1017/ewmxxx)

© .

All rights reserved. For Permissions, please email:

munity. As part of the common law, judges have significant discretion 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 a bilateral balancing model. Under this model, judges determine the proper standard by balancing the rights and interests of speakers and their targets. A desirable standard is one that offers targets sufficient protection without unduly chilling speakers from engaging in public debates.

Recent work sought to challenge the dominance of the bilateral model. Hemel and Porat (2020), contemporaneously with Arbel and Mungan (2020), argue that defamation law has broader, third-party effects. These papers develop a theoretical model where rational members of the audience assign higher credence to statements made in the presence of stricter defamation law, i.e., higher damages, lower burdens of proof, or more lower litigation costs. Audiences reason, in these models, that speakers would be less inclined to lie for fear of legal sanctions, and so it is rational to trust their statements more.² The converse is also true: weak defamation laws induce mistrust, leading audience members assign low credence to statements and to instead rely more heavily on their prior beliefs. As a result of the credibility effect, the strictness of defamation law is expected to have a direct effect on members of the public and affect their trust in speakers and the media.

If these theories are correct and defamation law creates a credibility effect, then this has 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 one seeks to diminish the perceived credibility of problematic sources, say tabloids or social media, deregulation could be helpful. More generally, the neglect of credibility effects means that modern standards are likely improperly balanced (Arbel & Mungan (2021)).

Why have courts failed to consider audience in their rulings? There are three main objections to the realism of the credibility effect. One is *informational*—Lay audiences are not likely to master the intricacies of defamation law. The second is *legal consciousness*: even if audiences are aware of defamation law, it may not be on their mind when they read the daily paper. The third, and most challenging one, is *cognitive*. There are a variety of studies that show that individuals deviate from the prescriptions of Bayesian decision theory—medical professionals’ base rate neglect in interpreting test results being a prominent example (Casscells et al., 1978). Such deviations are of special concern in the context of defamation law, because the law affects the speaker, not the listener. The audience is theorized to account for the speaker’s incentives in light of

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

2. The behavioral assumption that speakers would be reluctant to lie is of course debatable, but it is widely shared by courts and commentators. Notably, the standard account assumes a degree of legal awareness by lay speakers.

the law, but this accounting may implicate the fundamental attribution error (Ross, 1977). According to various studies, individuals systematically fail to consider the effects of external, situational incentives of individuals and instead attribute their actions to their internal disposition and character. There are also some doubts whether some adults possess the requisite theory of mind, the ability to reason from the viewpoint of another, that is required by the theory.

Given the policy importance of the credibility effect, empirical evidence is needed. To date, however, there has been little empirical work done on defamation law, and none that directly studies the credibility effect. This article seeks to offer the first step in this direction. Section 2 provides a brief legal background and a review of the extant evidence. Section 3 presents the primary study, a randomized vignette trial of the credibility effect with 200 participants. Respondents were presented with various factual statements made in four different media—TV news, newspaper, a blog, and social media—and were asked to evaluate the credibility of these statements. The scenarios were designed to mimic as closely as possible realistic conditions. The treatment consisted of the defamation law regime, with half of respondents assigned to an effective defamation law regime (where it is easy to prove defamation and recover damages) and half to an ineffective regime. The study establishes the existence of a clear credibility effect: the legal regime affects the trust ordinary folks place in information provided by the media. Notwithstanding the concerns with cognitive limitations, the assignment to an effective defamation regime resulted in an increase of 19% percent in the credibility of statements. A qualitative analysis of responses suggests that the effect is mediated by higher perceived source credibility, as predicted by theory.

The analysis presented in Section 3 invariably involves priming subjects to the existence of defamation law. While it establishes that lay individuals are capable of factoring the law into their analyses, it leaves open the question of the strength of this effect outside the lab. To that end, Section 4 reports a study designed to circumvent the priming of respondents. In this study, 200 respondents from both the USA and the UK were asked to evaluate the credibility of four news reports. The key design feature is that credibility assessments are elicited before defamation law is mentioned. After subjects provided their credibility assessments, they were asked to evaluate the 'actionability' of the statement. Actionability is a measure of the likelihood of the victim to win a lawsuit against the speaker if the statement is false. A lack of correlation between actionability and credibility will be consistent with the theory that defamation law does not impact credibility assessments.³ Instead, the study finds a strong positive association between actionability and credibility judgments made by lay individuals. The strength of this finding suggests that credibility assessments and perceptions of the legal regime are linked—within the inferential limitations from correlation analysis. This finding provides preliminary evidence that lay individuals may indeed factor legal constraints into their

3. Indeed, the existence of cognitive biases would lead us to expect a negative association between credibility and actionability. See discussion in Section 3

credibility assessments.

Section 4 briefly considers the vindication effect. A standard justification of defamation law is that it creates a venue for individuals to defend their name from malicious attacks. In this study, 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 attached a strong stigma to individuals who did not sue—but only when defamation law was strictly enforced. The stigma resulting from the credibility effect suggest the subtle nature of defamation law balances which is obscured by the dominant bilateral balancing model.

Section 5 concludes by discussing limitations and policy implications. A key observation is that lay individuals' are likely to be under-informed of the exact legal standard. This suggests that under certain conditions, the credibility effects found in the lab may be *stronger* outside the lab. Another relevant observation is that defamation law is highly salient, significantly more than many other areas of law. Defamation cases attract considerable press coverage and social media threads around news reports usually mention the possibility of a lawsuit. This suggests an above-average legal awareness of the law, notwithstanding misperceptions about its content.

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 make defamation law stricter by, e.g., 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 it might engender systemic fragility that can be exploited by adversarial parties. Interestingly, the discussion of misperceptions suggest that policymakers can achieve significant behavioral effects without changing the law—by simply educating the public about the legal standard. The stigma associated with strong defamation laws counsels in favor of moderation, or at least a more nuanced cost-benefit analysis. 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 labelling, and stolen valor. While striking the right balance can be difficult, it will be impossible to do so without accounting for credibility effects.

2. Background and Motivation

Defamation law regulates the dissemination of false statements that threaten their target's standing in the community. Under defamation law, a target of a (1) public statement that is (2) false and (3) harmful to one's 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, 2007).

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 determined that the legal standard should create a 'breathing space' for speakers.⁴ The concern with speakers also encompasses publishers, and economically oriented scholarship has emphasized how defamation law affects investigative journalism and the media (Garoupa, 1999a,b, Bar-Gill & Hamdani, 2003, Manoj & Refalo, 2007).

Responding to these competing concerns, the Supreme Court in *Curtis Publishing v. Butts*⁵ summarized the court's role in defamation cases in terms of a bilateral balancing model. The Court stated 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."

Until recently, little attention was given to the audience effects of defamation law. This omission is significant and curious. Significant because defamatory speech is only harmful if it is believed by audiences, thus it appears critical to account for the direction and magnitude of change in audience perceptions of the target. And curious because the marketplace of ideas, a key metaphor in First Amendment jurisprudence, is constituted mostly by audiences. Omitting audiences from the analysis creates a deep gap in our understanding of this marketplace. Recent work argued that accounting for audiences requires recognizing potential credibility effects, where stringent defamation standards would increase audience trust (Arbel & Mungan, 2019, Hemel & Porat, 2019). If true, the credibility hypothesis has a broad range of welfare and policy implications which are studied in Arbel & Mungan (2021).

As noted, there are three primary challenges the realism of the credibility effect. As the majority of audiences are lay individuals, it is unlikely that most have a good sense of the admittedly intricate rules of defamation law. Lay individuals may be uninformed about the law, and so will not be capable of correctly assessing its effect. Second, consumers may simply neglect defamation law's role when they make credibility assessments in daily life. A potential mugger may be aware of the punishment for armed robbery but may not think about it in the heat of the moment. Awareness, then, offers a second challenge.

Arguably the most central challenge is cognitive: the credibility hypothesis potentially exaggerates the cognitive ability and aptitude of lay individuals. There are a large number of studies in behavioral economics which show deviations from the prescriptions of rational choice theory. Adjusting credences based on source incentives requires a certain degree of rationality which lay individuals may lack. In particular, the credibility hypothesis assumes that individuals have a mature 'theory of mind,' which is necessary to model other people's thought processes. It is arguable, however, whether ordinary people possess the requisite level of theory of mind (Keysar, Lin, and Barr, 2003, but see Khilvati et al., 2019).

While the existence of the credibility effect of defamation law was never

4. *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46, 56 (1988)

5. 388 U.S. 130 (1967)

tested, there are some empirical studies in adjacent contexts which are inconsistent with it. Several studies show that audiences do not alter their credibility assessments when they learn that an advisor is advocating a self-serving position (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 found that consumers exhibit similar levels of trust in disease prevention claims (which are regulated by the FDA) as structure claims (which are not regulated), suggesting that FDA information regulation does not affect trust in statements (France and Bone, 2005).

Empirical work on defamation is limited. Cohen et al. (2001), find in a set of experiments that individuals discount biased defamatory publications, but believe that *others* will be influenced by them (the 'third-person effect'). Work in communication theory found that the level of sentence directness 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 reasoning, two hundred 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, ensuring that respondents are attentive, fit the recruiting criteria, have good standing in the platform, and are indeed unique individuals. By now, the choice of MTurk as a recruiting tool is well analyzed and—given quality controls—is becoming a standard recruiting tool in many areas of research.

The demographics of the sample, relative to the general U.S. population (in parenthesis), are: 44% female (50.8%), median age 35 (38), 77.5% white (60.4%), median household income in the \$40,000-59,000 range (\$57,652), and college degree or higher education 71% (30.9%). We thus find that the sample skews white and is significantly more educated. There are no theoretical predictions as to the effect of race, although education levels may be associated with a greater propensity to engage in higher-level abstraction and calculations. This could lead to the sample over-weighting the relevance of defamation law relative to the general population. Still, post-hoc analysis did not identify any significant (economically or statistically) correlations.

The treatment consisted of assigning respondents to different defamation law regimes. The challenge here is to make the governing law clear to respondents, and this could not be done without both informing them on the relevant law and priming them to its existence. The inferential limitations imposed by this structure are discussed later in the article, but the primary goal of this study should be understood as testing the existence of the credibility effect in light

of contradictory predictions—rather than offering a precise measurement of its magnitude.⁶

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

The standard bilateral model employed by the court neglects audiences, thus implicitly assuming that defamation law does not have important third party effects. On this account, we would expect there to be little effect to the group assignment on credibility assessments. This is particularly true if individuals are incapable—for a variety of cognitive limitations—to factor in the speakers' incentives and update on their basis. A competing hypothesis, provided by the theoretical work noted above, is that the level of defamation law would have a positive effect on credibility judgments. Thus, the main hypotheses 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% 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 participants 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 man against man, and the fourth was a firm against a woman. The names are fictitious and most of the images were computer generated and are not of real individuals. Each scenario used a different medium and the scenarios each involved allegations that were sufficiently serious to make defamation law potentially relevant. The scenarios are reproduced below.

News Report. The first task presented respondents with a scenario involving a published negative statement that alleged facts against a private individual. The scenario, presented in the form of a (researcher designed) scan of a news article, details the financial problems of a local firm. It mentions, without supporting evidence, that some of the financial problems are due to embezzlement by the company's accountant—Ms. Miranda Dewey.

6. This is another reason why the design does not include an untreated group. Note, however, that inclusion of such a group would not allow us to cleanly compare control and treatment, as the treatment involves more than just a change in the legal regime

Figure 1 News Clip



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

Social Media. Many individuals rely heavily on social media for the consumption of news. This raises several novel challenges, given the low barriers to entry for the reporting of facts on these platforms. Lacking curation and oversight, many view the reliability of social media information as low. The next scenario involved a posting on social media, described below. 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, rendering defamation law irrelevant.

Figure 2 Social media



Respondents were asked to rate, on a 0–100 scale, the likelihood that Patricia Sutton stole from her employer. Again, the opportunity to explain was given, alongside a rating of how likely she would be to win a lawsuit if the allegation were false.

TV Reporting. This scenario involved a report on TV, in the news section, involving an individual who is arguably a public figure.

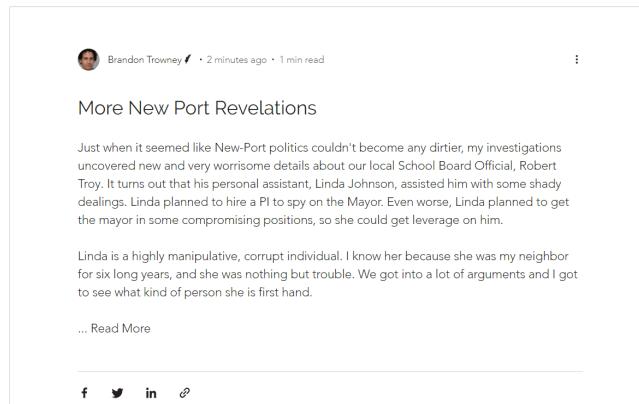
Figure 3 TV News



Subjects were told that they were watching the local news when they came upon this segment. They were told that Paul Stevens is a local businessman and that Jim Lane—the defamed—is a local contractor. The setup was 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.

Blog Post. The final scenario involved a blog post accusing a local woman, Linda Johnson, of attempting to collect compromising information about the local mayor to get leverage over him in an implied blackmail scenario. The same credibility assessment was repeated.

Figure 4 Blog Post



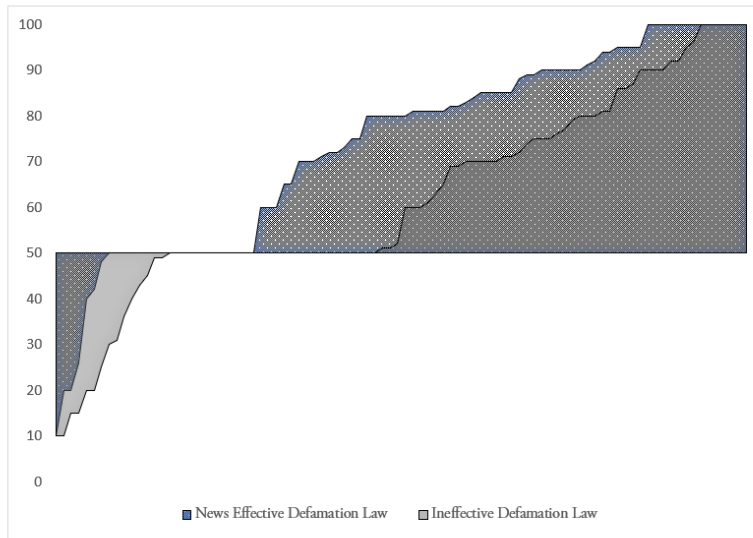
4. Findings

4.1 Scenario Analysis

Newspaper report: 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 most defamation law cases. In this case, a familiar scenario is presented, a news report makes factual allegations about a person, and the reader is left to decide how credible they find the report.

Figure 5 below shows how lay individuals ranked the credibility of the report between the groups. The Figure arranges the credibility assessments of participants in each group from the lowest assessment (0) to the highest (100). The x-axis is set at the point of 50—where the assertions were judged to be equally likely to be false as true.

Figure 5 Credibility of News Report by Defamation Law Regime



The responses by participants suggest that media reporting appears considerably more credible when defamation laws are strict. The effective defamation group felt very confident about the credibility of the report, with 62% of the sample giving it a score of 80 or above. For comparison, in the ineffective defamation group, the same percentage gave the story a ranking below 70. The mean credibility assessment in the effective defamation law group was 73.5, relative to 62.7 in the ineffective defamation law group—a difference of 18%. The difference was highly statistically significant.⁷

When asked about their reasoning, it was mostly respondents in the effective defamation law group who emphasized the role of defamation law. The following quotes are followed by the credibility assessment given by the respondent: “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)”.

Not all respondents mentioned defamation law and some made judgments based on other factors. It was remarkable, however, the level of confidence

7. News report $t(184) = 3.42, p < 0.01$

they pronounced in these judgments. For example, “The fact that Miranda is the company’s accountant led me to make that judgment. I feel she was in a position to harm the company and she could have used the low demand in iron to cover her embezzling (88)” or “I pretty much made my judgement based on the newspaper. It is likely that she stole money from the company (72)”.

By contrast, the responses in the ineffective defamation law group were more agnostic and, rather than drawing a negative inference from the weakness of defamation law, simply did not mention it as often. One respondent said “The reason the company is doing poorly is due to lack of resources not because of embezzlement (20)” and another stated “I really would have no idea if this is true without more context or information so to me it is just a toss up without more information (50)”. Still, a small number of respondents did 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)”. In the absence of direction, respondents emphasized different credibility cues. Some expressed reliance on journalistic integrity, “I picked 75% likely because if a newspaper reports it, they usually try to get the facts right, but sometimes are wrong (75)”. Others focused on the phrasing “It’s a newspaper, which suggests it should be credible, but then it’s written like the embezzlement is no big deal because it only goes on for a sentence and then switches back to talking about something else. This seems fake (10)”. Overall, news reporting was judged as considerably more reliable in the presence of effective defamation law. Respondents seemed more likely to draw a positive inference from defamation law than a negative inference from its absence.

Social Media: Individuals consume a great deal of information through social media. While there is sometimes personal familiarity with the individuals involved in posting, it is not uncommon to encounter posts by unfamiliar individuals. The potential harm of social media posts depends on their credibility, and so lawyers sometimes make the argument that a post is not actionable because it was posted on social media—and that such posts are naturally processed as less credible by their audiences. Figure 6 reports the credibility assessments of individuals on the same 0–100 scale as above.

Figure 6 Credibility of Social Media by Defamation Law Regime

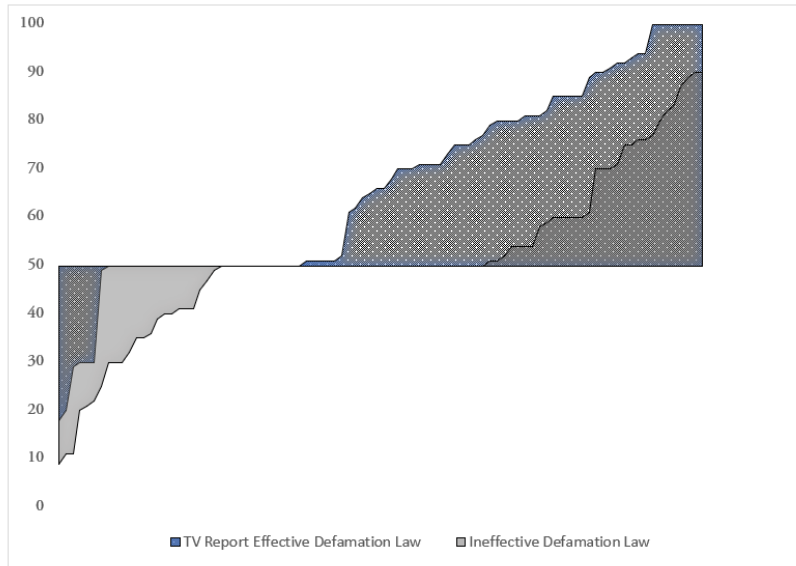


Figure 6 indicates that respondents were more likely to believe postings on social media when defamation law was strict. The level of confidence was not as strong as in news reporting (in fact, the level of confidence in social media was the lowest across all conditions); still, 36% ranked the statement at 80 or above in the effective defamation regime, relative to only 16% in the ineffective regime. The mean credibility evaluation stood at 59.2 in the effective defamation law regime, relative to 51.8 in the ineffective regime—a 14% difference. The differences were statistically significant.⁸

The reasoning provided for respondents in the effective defamation law group emphasized the negative consequences that will be suffered by the company for lying. For at least one respondent, defamation law was the deciding factor: “I am leaning towards the company being right as they would have the most to lose in a defamation case (60)”. One important factor was the company’s sophistication, as noted by one respondent: “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)”. On the other hand, the company’s sophistication also resulted in higher expectations, with various respondents commenting on the unprofessional attitude of accusing ex-employees on Facebook. “I don’t think it’s likely that she stole or the company wouldn’t be so unprofessional as to claim that on a public platform. Instead, they would’ve taken legal action. I think it’s defamation (30)”. As before, some respondents did not mention defamation law, despite having made a confident assessment:

8. Social Media $t(184) = 2.27, p < 0.05$

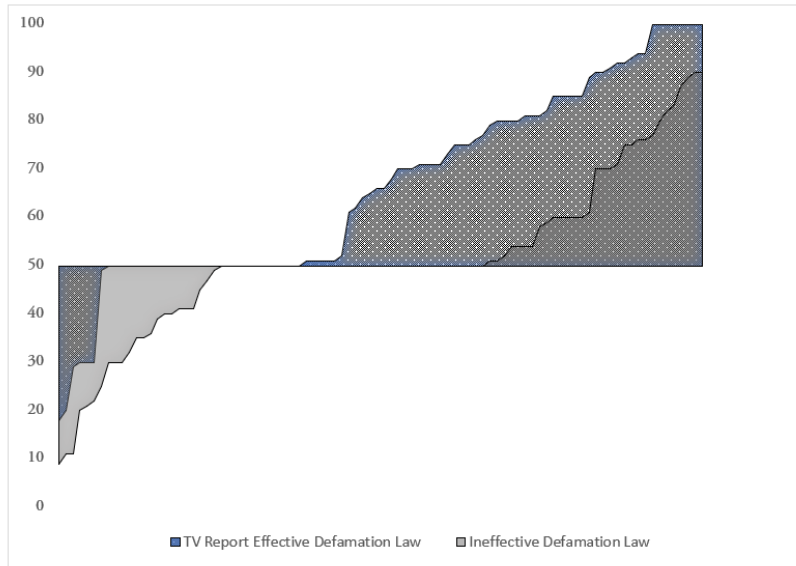
“Patricia’s statement is a subjective opinion about her employer. Also, it seems that the company was justified in firing her (72)”.

Interestingly, respondents factored in not just the likelihood of winning a lawsuit, but also of filing one. One respondent noted that “Patricia doesn’t have a lot of money to pay lawyers (80)”. Likewise, another respondent considered the disutility of suing Patricia, “FRI Photography has much more to lose than an individual person, here. 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 (85)”.

In contrast, in the ineffective defamation law group, respondents emphasized the lack of evidence. “it’s all just hear-say from both sides (50)”. They relied on a variety of cues, “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)”; “they could have deleted the comment or blocked them if it was true (32)”. One respondent relied on reputational concerns “even in this state, I don’t think a company would risk the negative PR of lying about an employee (60)”. Some expressed platform skepticism “It’s a facebook post, anything you read on social media is about as accurate as flipping a coin (50)”. A small number of respondents relied on the weakness of defamation law for adverse inferences, “I can’t know the truth when I have such little context. And the poor defamation laws just add to my uncertainty (55)”. In sum, social media postings were not deemed as being very credible—but were seen as considerably more credible in the presence of effective defamation laws.

TV News: TV reporting is a markedly different industry than print news, and so it raises some special considerations. TV reports involve a dynamic visual experience that is lacking in the consumption of print media, and while it was difficult to recreate the full experience here, it was suggested by the specific use of imagery. The channel itself was unfamiliar—“NP Local News”. The scenario is meant to evoke general perceptions of TV reporting, without the specific polarizing valence attached to channels such as Fox News or MSNBC. As in the other conditions, respondents were asked to rank the credibility of allegations made by an individual appearing on TV. Figure 7 below marks their responses in the two groups.

Figure 7 Credibility of TV News by Defamation Law Regime



Respondents found TV reporting to be considerably more credible when defamation law was effective than when it was ineffective. In fact, only 16% of respondents ranked the report at 80 or above in the ineffective condition, relative to 45% in the effective defamation law condition. This shows a large impact of defamation law on TV reporting—with the mean being 66.1 in the effective defamation regime relative to 52.8 in the ineffective regime. The difference of 25.3% between the two was the largest across the scenarios, suggesting that defamation law may play an especially important role in this context. The difference between the means was highly statistically significant.⁹

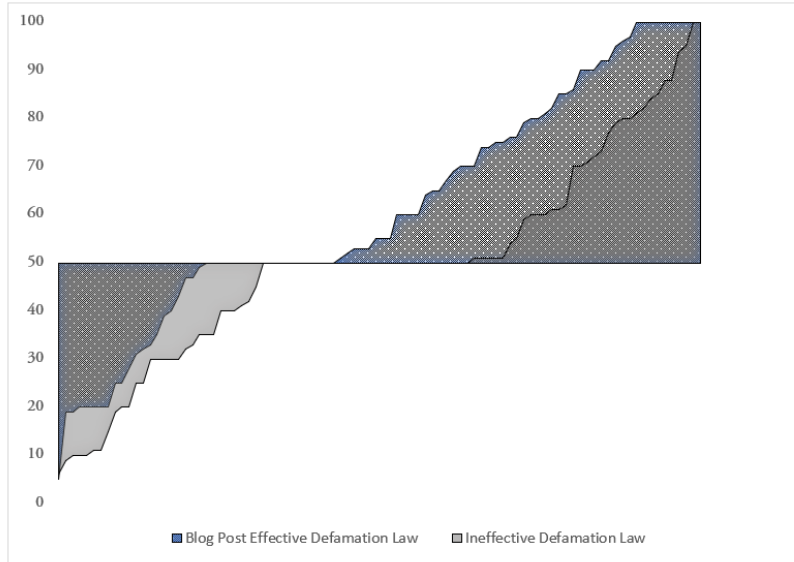
The reasoning provided by respondents in the effective defamation law group repeated earlier statements about the deterrent effect of defamation law. This time, respondents also emphasized the sophistication of the news organization, which made it aware of the content of defamation law: “I think in this state the news for sure knows the law and does not risk stating anything before they fact checked it (100)”. The stations’ status also lent credibility to the report: “I would imagine that any TV station that has established itself as trustworthy probably has pretty high standards for reporting a fact-checking, so I find it more likely than not that this accusation is true (90)”. One respondent explained that she or he gave a lower credibility assessment (relative to the previous scenarios) because “as a private citizen, Mr. Stevens may not be aware of the stringent penalties for making false claims (85)”. Others focused on the public nature of the accusation: “Paul Stevens must be pretty certain of the allegations, and even have proof, to make that claim so publicly (75)”.

9. TV News $t(184) = 4.75, p < 0.01$

In the ineffective defamation group, respondents' reasoning emphasized again the weakness of defamation laws. "The laws do not prevent people from making false accusations so there is no reasonable way to tell if Jim did what he was accused of just by Paul's statement (50)". Instead, respondents focused on the paucity of evidence "This is just a case of he said she said, and there's no reason given for proof (39)", and searched for likely motives "The businessman is most likely wanting to get revenge for losing a contract to the contractor. (25)"; verbal cues "The statements seem speculative and opinionated, so I do not think the claim is true. It does not read like something that is well researched. It sounds like someone's personal opinion (30)"; or simply defaulted to their prior "I think Jim Lane probably did pay a bribe, because that's just what contractors do (70)" and "News companies and stations often falsely or inaccurately report information like this (20)". Consistently, there was a recurrent theme of agnosticism "Again, since everyone is fair game thanks to the law, not only is it improbable that Jim will win a defamation lawsuit, it's also likely that what he was accused of was false. Again, though, it's possible he did pay a bribe, so I'm going with it being a 50/50 shot (50)". Interestingly, even when respondents expressed positive trust in TV news, their quantitative assessment was generally lower: "It looks like it made television, which may be fact checked thoroughly (41)"; "It's on the news so there has to be some credit (61)". Overall, the credibility of TV reporting was highly influenced by defamation law.

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 appeared sufficiently interesting to merit an investigation. In the blog post, a local person accuses a local woman of various misdeeds. Respondents were asked to assess the credibility of the allegations and their responses are recorded in Figure 8 below.

Figure 8 Credibility of a Blog Post by Defamation Law Regime



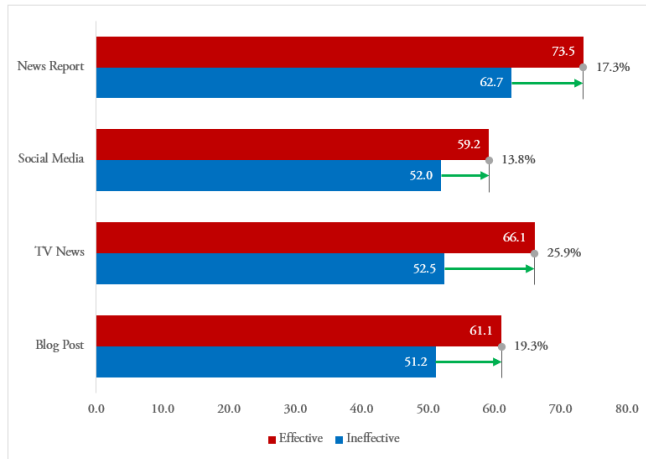
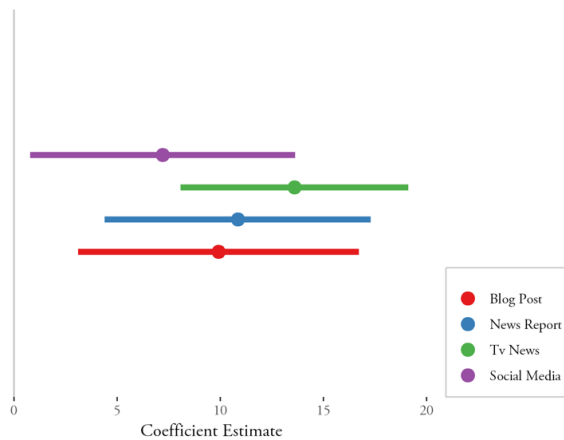
As with other cases, respondents again ranked the blog post as significantly more credible when defamation law was effective. In the presence of effective defamation law, the percent of ranking at 40 or below was 18%. In the absence of defamation law, respondents ranked the blog post as being highly unreliable, with 28% of respondents ranking it at 40 or below (the worst of all conditions). The mean credibility of the blog post was 61.1 with effective defamation laws, relative to 50.9 with ineffective defamation laws— a 20% difference. The differences were highly statistically significant.¹⁰

To summarize the effect of defamation law on the assessment of credibility in each scenario, Figure 9 below offers a side-by-side comparison of the means under each condition. The move from a state with ineffective defamation laws to a state with effective defamation laws causes a marked difference of 14–25% in the evaluation of the credibility of statements. As noted, all the differences are highly statistically significant, as illustrated by the representation of the slope coefficients in Figure 10 below.

Interestingly, a post-hoc analysis did not find any statistically significant interaction effects between the scenarios along the effectiveness of defamation law. This is of interest because it is plausible that defamation law would have different credibility effects in different contexts, but the analysis did not locate any such interactions. While this may be an issue of power, it is also possible that defamation law works similarly across contexts.

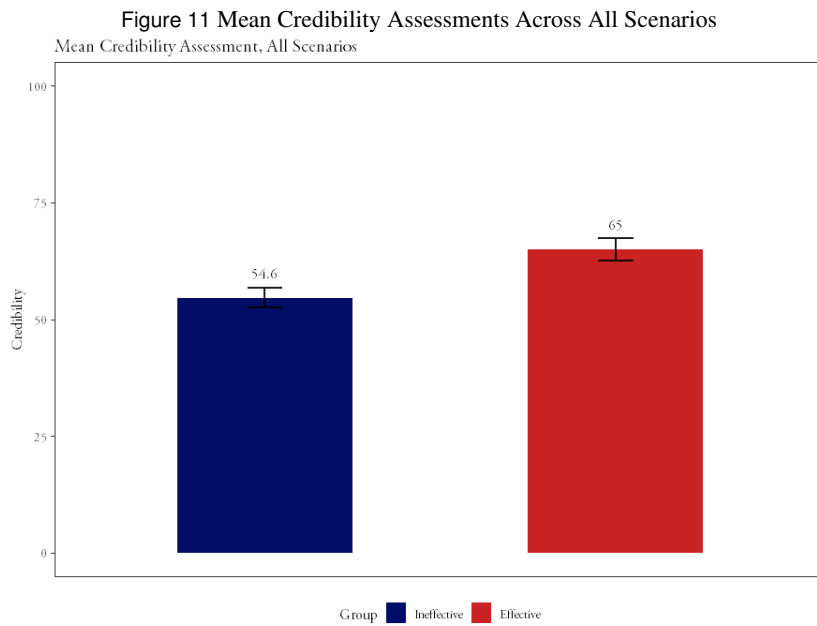
10. Blog post $t(184) = 2.95, p < 0.01$

Figure 9 Mean Credibility Differences by Defamation Law Regime

Figure 10 Slope Coefficients within a 95% Confidence Interval
Scenario Specific Credibility Effects of Defamation Law

4.2 Group Analysis

The discussion above recorded the existence of a credibility effect for each scenario. Looking at the findings as a whole, Figure 11 below compares the mean credibility assessment in each group. The mean credibility assessment was 10.4 basis points higher in the effective defamation group (an increase of 19% percent). The effect is statistically and practically significant.¹¹



Controlling for a variety of demographic factors, the regression table below reports the effects of defamation law on credibility assessments across all conditions. There were no statistically significant effects due to age, race, gender, income, or education.

11. $t(742) = -6.3, p < 0.001$

Figure 12 Regression Analysis
Results

	<i>Dependent variable:</i>						
	Credibility						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effective Defamation Law	10.383*** (1.648)	10.455*** (1.651)	10.405*** (1.662)	10.173*** (1.651)	10.449*** (1.651)	10.076*** (1.670)	9.948*** (1.698)
Age		0.060 (0.070)					0.051 (0.073)
Asian			-1.778 (2.487)				0.260 (2.614)
Black or African American			-0.265 (3.880)				1.231 (3.943)
Other race			-2.526 (3.531)				-1.092 (3.664)
Household Income over 60,000				-2.755 (1.659)			-3.195 (1.712)
College or higher education					-1.389 (1.829)		-1.715 (1.866)
Male						0.103 (1.680)	0.031 (1.712)
Other gender						-14.996 (8.092)	-16.148 (8.410)
Constant	54.614*** (1.159)	52.269*** (2.977)	54.995*** (1.243)	55.963*** (1.414)	55.575*** (1.716)	54.871*** (1.566)	55.669*** (3.785)
Observations	744	744	744	744	744	744	744
R ²	0.051	0.052	0.052	0.054	0.052	0.055	0.062
Adjusted R ²	0.049	0.049	0.047	0.052	0.049	0.051	0.050

Note: $p^* < 0.05, p^{**} < 0.01, p^{***} < 0.001$

4.3 Summary

Taken together, these findings show the causal effect of defamation law on credibility assessments. Moving from an ineffective defamation regime to an effective one leads to a 19% change in the credibility assessments of respondents. In the abstract, the effect size is moderate, although in specific 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).

In each separate medium (Newspaper, social media, TV news, and blog) defamation law had a significant effect on the perceived credibility of the statements. People exhibited higher trust in the media when defamation law was made more effective.

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.

There are a number of limitations that affect the interpretation of these findings, summarized in the Discussion section. The most important one is the fact that respondents in this study were *informed* about the content of defamation law and, consequently, may have *primed* them to think about the law. The next study seeks to address this concern.

5. Credibility Effects in Ordinary Judgments

The previous study demonstrated that lay individuals are capable of, and are in fact, responding to the legal regime, and that the effect size in the lab is moderate. These findings naturally leads to the question of the size of the effect in everyday life, where individuals are often under- or misinformed about the law, and are not primed to consider it.

Unfortunately, measuring the effects of defamation law in real life settings is confounded by a number of factors. One naive approach is to measure the difference in perceived statement credibility between individuals from different jurisdictions, and interpreting the differences as resulting from the differences in legal regimes. This approach, however, would not provide a satisfactory answer, as there are many other differences between jurisdictions and the individuals who reside in them. Perhaps most critically, individuals within a jurisdiction have different perceptions of the legal standard, and even on average their perceptions may not match the true standard. Any attempt to correct misperceptions through disclosure of the true legal regime would involve informing and priming respondents.

Recognizing these challenges, the present study takes a different approach. Instead of looking for the effect of the legal regime on credibility assessments, it focuses on the effects of the *perceptions* of the legal regime. By doing so, the study can measure effects without informing or priming individuals. The key feature of the study is that credibility assessments are made first and legal perceptions second, seeking to determine the existence of a possible association

between credibility judgments and perceptions of the legal standard.

5.1 Research Design and Methodology

Two variables inform the analysis. The assessed credibility of a statement—how likely the individual thinks that the assertion made is factually correct—and the perceived ‘actionability’ of the statement. Actionability is defined here as a measure of the probability that if the plaintiff were to bring a lawsuit against a false assertion, then the plaintiff would prevail in litigation. From the lawyer’s perspective, statements of opinion have low actionability, whereas statements of fact have high actionability because the law generally exempts opinions from the scope of defamation law. While actionability may have a precise sense in the mind of the lawyer, our interest here is in lay perceptions of actionability.

On reflection, three hypotheses present themselves regarding the relationship between the credibility of statements and their actionability.

(1) H_0 The Independence Hypothesis. Under this hypothesis, credibility assessments are independent of actionability perceptions. The basic version is that lay individuals tend to lack critical thinking skills and prefer to believe statements that accord with their priors, group affiliation, or what may seem fashionable and high-status within their social group. A richer version posits that even if individuals are motivated to search for truth, the cues they rely on are extra-legal. Credibility judgments may feature a complex set of textual, contextual, or situational cues, leaving little room for legal factors. Defamation law, in this account, may be too remote to have appreciable impact. Moreover, reasoning on the basis of the legal standard requires a high level of abstraction and mature theory of mind, and it may be that many individuals lack these cognitive skills.

(2) H_1 The Bias Hypothesis. Individuals may be over confident in their assessments of the credibility of statements. After expressing a belief in a statement, individuals may be reluctant to admit the possibility that a judge would rule that it is false. This conflict creates a cognitive dissonance, which is likely to result in deflated perceptions of actionability for individuals who believe in the statements (and conversely for those who do not believe in them). Overall, if some version of this is true, we would expect a negative relationship between credibility and actionability judgments.

(3) H_2 The Credibility Hypothesis. As before, this hypothesis posits that credibility and actionability are intimately linked. A speaker incurs a risk when he makes a false statement, and the risk increases the more actionable the statement is. Thus, if a speaker decides to make a statement notwithstanding the risk, it is a signal that the speaker believes the allegation to be true. In other words, if a speaker “puts his money where his mouth is,” it is more likely that what he says is true. Thus, we would expect audiences to find statements that are more actionable to appear more credible.

To test these hypotheses, two hundred participants from the United States and the United Kingdom (101 US, 99 UK) were recruited. The demographics

of the American sample, relative to the general U.S. population (in parenthesis), are 39.6% female (50.8%), median age 32 (38), 67.3% white (60.4%), median household income of \$72,000 (\$57,652), and college degree or higher education 76.2% (39.1%). The most significant skew was education, and the sample was otherwise somewhat wealthier, male, younger, and white than the general population. With the exception of education, none of these factors is theoretically expected to affect the findings. In the UK sample, the demographics were 67.7% female (50.5%), median age 32 (40.5), 97% white (87.2%), median household income of \$46,900 (\$40,484), and college degree or higher education 59.6% (38.6%). The most significant skew was again education, although to a lesser extent than the US sample. This sample was somewhat younger, whiter, and wealthier than the general population. There was also a gender skew in this sample, but this time females were over-represented.

The participants were asked a few demographic questions and then moved to a page where they were presented with an attention check question, followed by four different scenarios. 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 this presentation.¹² For concreteness, one of the four scenarios is reproduced below:¹³

The design of these scenarios was made with a few principles in mind. The situations they describe are familiar and involve diverse scenarios. In all of them, the allegations are factually ambiguous and judgments have to be made based on limited information. Even though defamation law is not mentioned, the scenarios would be handled differently under defamation law—most clearly, the difference between public and private figures would make the first statement less actionable than the second, at least under US law.

Scenario: News, Public Figure

You read the local paper. You happen to read a story by a local reporter which contains the following line: "Jennifer Davidson was caught embezzling in her last position and had to step down" You don't know her well, but from what you know, she is a local politician, currently running for office.

How likely is the statement about Jennifer to be true?

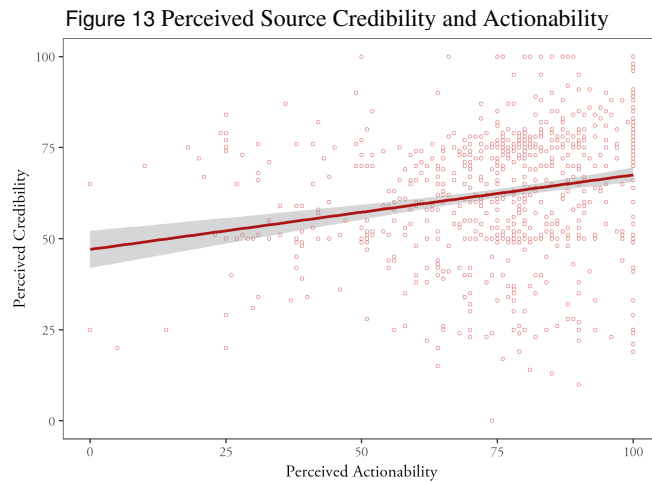
After making credibility assessments, respondents were moved to the next stage and were unable to go back and revise their responses. In the second stage, they were told that the allegation is false and were asked how likely the plaintiff is to prevail in litigation assuming he or she has reasonable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

12. Per IRB standards, the invitation to the survey identified the lead research 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

13. see Appendix A for the remaining scenarios

5.2 Findings

The responses were aggregated into 800 response pairs (credibility and actionability). The analysis of the responses, graphed in Figure 13 below, shows that individuals made credibility assessments that were strongly correlated with their assessments of the legal exposure of the statement, i.e., its actionability. Across all scenarios, there was a positive correlation of $r = 0.21$ between the perceived credibility of a statement and its perceived actionability. The relationship was highly statistically significant.¹⁴



The table below reports the full results of a regression analysis. The association between actionability and credibility holds while controlling for other demographic factors. The most surprising finding is that American respondents were *more* likely to find the statements credible. I hoped that inclusion of respondents from two jurisdictions with different defamation regimes would allow us to identify the existence or absence of credibility effects. This did not pan out, because whatever real differences exist between the legal regimes, they did not manifest themselves in the perceptions of respondents. Indeed, Americans were slightly (but statistically insignificantly) more likely to view statements as actionable ($\mu_{UK} = 75, \mu_{USA} = 76$). There may be a number of reasons for those similar perceptions, and it is independently interesting that Americans seem to place more trust in publications than their English peers—but a full investigation of these questions will await another time.

14. $t(798) = 6.14, p < 0.01$

Figure 14. Analysis of Credibility Perceptions as a Function of Actionability Perceptions

Results								
	Dependent variable:							
	Credibility							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actionability	0.205*** (0.033)	0.203*** (0.033)	0.201*** (0.033)	0.207*** (0.033)	0.206*** (0.033)	0.199*** (0.033)	0.202*** (0.033)	0.192*** (0.033)
Male		4.554*** (1.245)						3.425** (1.314)
Other gender		11.309* (4.417)						8.468 (4.488)
USA			5.821*** (1.223)					5.088*** (1.420)
Asian				0.801 (2.444)				-1.234 (2.534)
Black or african american				1.036 (2.132)				-3.239 (2.256)
Other race or ethnicity				-3.559 (5.111)				-7.123 (5.094)
Household income over 60,000					0.00002 (0.00001)			-0.00000 (0.00001)
Bachelor's degree or higher						2.825* (1.328)		2.085 (1.420)
Age							0.072 (0.057)	0.041 (0.059)
Constant	47.087*** (2.591)	44.965*** (2.617)	44.462*** (2.615)	46.804*** (2.641)	45.935*** (2.758)	45.561*** (2.683)	44.940*** (3.107)	41.936*** (3.272)
Observations	800	800	800	800	780	800	800	780
R ²	0.045	0.066	0.072	0.046	0.049	0.051	0.047	0.088
Adjusted R ²	0.044	0.063	0.069	0.041	0.047	0.048	0.045	0.076

Note: *p<0.05; **p<0.01; ***p<0.001

Beyond the jurisdictional effect, education and gender also play a role. Inasmuch as education is a proxy for sophistication, a plausible explanation is that sophisticated individuals are better positioned to make more complex credibility assessments. The effect of gender is unclear, and should await future exploration.

Broken down by scenario, the association was highly statistically significant in three of the four scenarios.¹⁵ While the correlation ranged between 0.18–0.27 for the three scenarios, it was close to zero in the case of a news report of a public figure. This last finding was unexpected, but may be explained on the basis of the relative immunity given in the US to the media on issues of public concern.

5.3 Discussion

This study shows a statistically significant correlation between lay perceptions of credibility and actionability. The interpretation should proceed with

15. Bulletin Board $r = 0.2$, $t(198) = 2.88$, $p < 0.01$, Company email $r = 0.27$, $t(198) = 3.93$, $p < 0.01$, News, public figure, $r = 0.05$, $t(198) = 0.75$, $p = 0.45$, News, private figure $r = 0.18$, $t(198) = 0.75$, $p = 0.01$

the recognition that the findings show correlation, not causality. With that in mind, the association is moderate, suggesting that to the extent that actionability influences perceptions of credibility, it is not a dominant factor. This is expected given the plethora of alternative credibility cues besides defamation law, such as the plausibility of the claims or the medium. At the same time, there is some reason to believe that the true effect size is in some sense much larger. To the extent that some respondents are overconfident in their credibility assessments, they may be biased to evaluate the statements as *less* actionable, to avoid a cognitive dissonance between statements they believe to be true and the possibility that a court would rule that they are not.

Independent of the exact effect size, the primary insight here is that respondents assigned higher credibility to statements that they thought were more actionable. In other words, respondents believed speakers more when the speakers stood to lose from making false allegations. Importantly, respondents were not primed to the existence of defamation law, and so this finding complements the findings of the previous study.

The findings of this study are consistent with H_2 , the credibility hypothesis, subject to the caveat that this study measures correlation not causation. Importantly, the findings are inconsistent with both H_0 and H_1 , i.e., the independence and bias hypotheses. The finding of this inconsistency has significant legal and practical importance. The independence hypothesis underlies a significant body of jurisprudence developed by the Supreme Court over the last six decades, which is manifested in the bilateral balancing model.

One challenge to the interpretation of the finding is the possibility of acquiescence bias in responses (Bentler, Jackson, Messick, 1971). A positive relationship may result from subjects unthinkingly responding in the affirmative to questions. A comparison of the distribution of answers mitigates this concern. On actionability judgments, the center of the distribution mass is around 75 in all scenarios; but in credibility assessments, the center of the mass was around 50 in two scenarios and 75 in the other two (the shape of the distribution is also different). In other words, there was variability between responses that mitigates the concern of acquiescence bias.

Another important aspect of this study is that it measures *perceptions* of defamation law, and the diversity of opinions clearly shows that misconceptions abound. The existence of misconceptions is important to policy design, because it suggests that policymakers can affect behavior without necessarily changing the legal standard. They can affect behavior by informing (or, more cynically, un-informing) the public about the prevailing legal regime.

6. Credibility Effects and Vindication

The Supreme Court stated 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.”¹⁶ Two leading scholars, Professors John Goldberg and Benjamin

16. *Milkovich v. Lorain Journal Co.*, 497 U.S. 1, 23 (1990). (quoting *Rosenblatt v. Baer*, 383 U.S. 75, 93 (1966) (Stewart, J., concurring)).

Zipursky, explain that a justification for defamation law is the fact that it provides “right to vindication in the face of ... deliberate and malicious attacks by others.” (2020).

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

There is, however, a competing theory. The vindication theory expects the public to update its perceptions of the victim when a redeeming judgment comes to light. The same public should also be able to observe the failure of victim to bring suit (a fact which the speaker may strategically advertise). This creates a vicious dynamic from the victim perspective. Failure to bring a lawsuit marks one with an even deeper stigma—although there may be many legitimate reasons why a lawsuit is not brought. The stronger defamation laws are, the greater the stigma attached to *failing* to bring a lawsuit.

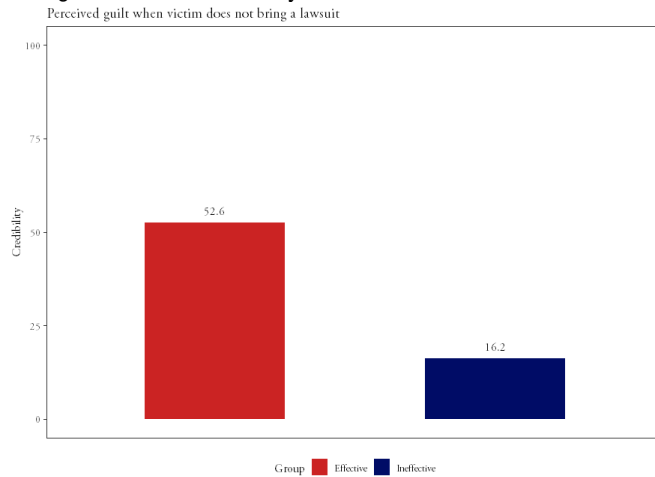
To explore the ‘anti’ vindication effect, the randomized controlled trial reported above ($n = 186$) included a vindication question. In the blog post scenario, where Linda Johnson was accused by a local blogger of attempted blackmail of the local mayor, respondents were asked to evaluate the credibility of the claim. They were then asked to rate, on a 0–100 scale, the following proposition: If Linda Johnson fails to sue the blogger, that means she is guilty of what she is accused of. analyzed.¹⁷

Figure 15 reports the findings of the analysis. The mean perception of the victim’s guilt from failure to bring a lawsuit in the effective defamation law group was 52.59, relative to only 16.2 in the ineffective group. The difference was highly statistically significant.¹⁸ People draw a much more negative inference about the subject’s guilt from failure to sue when defamation laws are strong. In other words, when the victim stands to gain a lot from a lawsuit but chooses not to file, this is perceived as an implicit admission of guilt. Defamation law thus casts those who do not sue in a negative light.

17. Attention check questions involved basic reading comprehension questions and inattentive or suspected responses were removed. Including these responses does not qualitatively alter the results.

18. $t(184) = 7.98, p < 0.01$

Figure 15 Mean Credibility Assessments Across All Scenarios



This finding highlights some of the negative side effects of the credibility effect. It marks some innocent victims with suspicion and puts pressure to litigate on people who would otherwise not sue. In the overall analysis, it is not clear what the net effect of these offsetting effects is, nor is it clear that the metric should be net effects. Society may worry about credibility side-effects if they mainly affect weaker parties who are less likely to bring suit, notwithstanding any positive effects of defamation law on stronger parties. This conclusion should be of special interest to civil recourse theorists. The ability to bring suit is generally theorized to advance victims' well-being by expanding their choice set (alongside other societal goods). The credibility side-effect raises the concern that there may be 'too much' recourse. In contexts like defamation law, the presence of easier access to the courts can pressure people not interested in litigation into lawsuits, and cast those less likely to litigate in a negative light.

7. Conclusion: Limitations and Policy Implications

7.1 Credibility Effects Outside the Lab

What is the size of the credibility effect in everyday life? The studies show that individuals adjust their credibility assessments based on defamation law. Even when not primed, credibility assessments are tied to the perception of the legal regime. A few considerations shed light on the relevance of credibility effects to everyday life.

One part of the question turns on legal awareness. Clearly, individuals cannot be expected to master the intricacies of defamation law doctrine. Indeed, many lawyers would find it difficult to articulate the doctrine. At the same time, misperceptions are not equivalent to ignorance or irrelevance. Consider that lay individuals are also unfamiliar with the legal standard that governs the enforceability of contracts, but rather than assuming that contracts are un-

enforceable, lay individuals tend to *exaggerate* the enforceability of contracts (Willikinson-Ryan and Hoffman, 2015, Sommers and Furth-Mazkin, 2021). Contract scholars worry that this bias leads lay individuals to comply with unenforceable agreements, avoid meritorious lawsuits, and attach exaggerated moral significance to contractual terms. Similarly, a survey found that 43% of consumer erroneously believed that only drugs that were completely safe could be advertised to consumers (Bell, Kravtiz, and Wilkes 1999). This belief, presumably, led consumers to purchase unsafe drugs. Thus, ignorance of the *specifics* of the law can lead to an exaggerated sense of its breadth, rather than an assumption that it is irrelevant. Because it is legal perceptions, not the law on the books, that affect behavior, it is an important task for future work to measure the direction of lay misperceptions.

A broader consideration is that modern jurisprudence *already* assumes awareness and consideration of defamation law by lay individuals. The bilateral model is premised on the concern that any change to the legal norm might deter ordinary people from participating in public discourse. If that were not true, the First Amendment would lose much of its vigor in this domain.

This assumption is not unreasonable. In almost all areas of law, the design of the legal standard incorporates behavioral responses—traffic regulation, tax code, criminal law, etc—which naturally assume some public awareness of the legal standard. To the extent lay individuals are aware of any law, defamation law enjoys high salience. Defamation lawsuits are often dramatic and attract considerable media coverage. Based on an anecdotal survey, many threads on social media related to news reporting tend to involve at least one commentator who mentions the possibility that a lawsuit will be brought if the story proves false. Thus, it is likely that the public has general awareness of defamation law, although the specifics must await future surveys.

7.2 Limitations

Beyond the limitations mentioned so far, a few other limitations constrain the interpretation of this study. The samples are not representative and skew more educated than the general population, among other differences. The second study showed that education was linked to a greater association between credibility and actionability perceptions, and so 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. As the challenges to the theory argue, consideration of speaker incentives requires a certain degree of sophistication. Thus, we would expect an education effect if the credibility effect is true.

Another issue involves the scenarios themselves. A routine concern in vignette studies is that the context-rich fact pattern might elicit responses to factors that are unrelated to the research question. E.g., people may perceive statements against women to be both more credible and more actionable. The use of four different scenarios in each study with variation in contextual factors involved (such as speaker's gender) mitigates some of these concerns, but cannot

eliminate them. The contrary concern is also relevant. Real life scenarios may involve more credibility cues, and would at least permit individuals to search for more information. The findings, then, should be interpreted as reflecting scenarios where 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 lax defamation law and another with a strict one. As a result, size of the effect (as distinguished from its existence) should be interpreted with some caution.

Notwithstanding these limitations, the present paper provides the first set of evidence on an important social question on which we lack key evidence.

7.3 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 policymakers. The existence of a credibility effect suggests a policy lever that is within policy-maker's control: defamation law. Increasing media and speaker 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.

The credibility effect also suggests the limits on defamation law. Reputational harms only come from people believing derogatory statements about others. If defamation laws are set too strict, they can counter-intuitively amplify the harm of false statements due to the enhanced trust they create (even if they reduce the overall volume of false speech). The finding that credibility effects also create stigma for those who do not file a lawsuit further reinforces the concern with strict defamation laws. These considerations suggest an important trade-off between credibility of speech and harm from statements. The welfare implications are investigated by Arbel & Mungan (2021), who provide an argument for moderation and erring on the side of free speech.

Doctrinally, the credibility effect suggests that the bilateral balancing model employed by courts fails to capture the true impact of the legal standard. Courts are balancing two factors when they should, in fact, balance three. In the future, courts should be attentive to credibility effects, and should make decisions that factor the pre-existing and desired level of trust within a given domain. Other things being equal, consideration of credibility effects can tip the scales in the context of social media in one direction, and printed news in the other.

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 the fact of 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 suggest care and development of better empirical understanding of audience dynamics.

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. Across this broad range of diverse contexts, we have neglected to consider credibility effects, and how regulation of falsities can be too strict for its own good.

Future work might be able to identify natural experiments that would allow us to discern the causal effect of changes to defamation law on measures of public trust. It will also be productive to learn about the type of misperceptions ordinary folks have about the law: do they over-estimate the scope of media liability? Do they underestimate liability on social media? Such measurements could inform future reform attempts.

8. Appendix A Scenarios from the Primary Study

Scenario: News, Private Figure

The same story you just read also contains the following line: "Miranda Dewey embezzled funds in her previous position and had to step down." You don't know Miranda well, but from what you know, she is a local accountant who mostly keeps to herself.

How likely is the statement about Miranda to be true?

Scenario: Email, Hiring

You work for a large company, and you are part of the HR department. You are now screening candidates for a position, and are considering a candidate called Dennis Wombly. An email goes to a number of people in the company, asking them if they know Dennis. In response, you see the following email from Matt. Matt works for a different department and you don't know him well. "I know Dennis very well because we worked at Dailysoft at around the same time and he was my shift manager. During his time there, Dennis went on a couple of business trips for the company. He took his wife with him and added her expenses as his own, which goes against company standards. A BAD hire"

How likely is the allegation against Dennis to be true?

Scenario: Email, Hiring

Scenario: Offline You enter a local coffee shop. As you wait for your order, you check out the bulletin board inside. You see a handwritten note that says: "Local attorney Ms. Donna Thomson is a SCAMMER. She took \$\$\$ she wasn't supposed to take and used it to pay herself. That's money that was supposed to be our DOWNPAYMENT. Better watch out. –Linda Paterson."

How likely is the allegation against Donna to be true?

References

1. Arbel, Yonathan A. and Murat Mungan. 2019. The Uneasy Case for Expanding Defamation Law. *Alabama Law Review* 1-999
2. Bentler, P. M., D.N. Jackson, S. Messick. 1971. Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin* 76, 186–204..
3. Cain, DM. M. Banker. Do Conflict of Interest Disclosures Facilitate Public Trust? *AMA J Ethics*. 2020 Mar 1;22(3):E232-238.
4. Casscells, W., A. Schoenberger, & T.B. Graboys. 1978. Interpretation by Physicians of Clinical Laboratory Results. *New England Journal of Medicine*, 299(18), 999–1001
5. France, Karen Russo and Paula Fitzgerald Bone. 2005. Policy Makers'

Paradigms and Evidence from Consumer Interpretations of Dietary Supplement Labels, 39 *The Journal of Consumer Affairs* 27, 42–45.

6. Hemel, Daniel and Ariel Porat. 2019. Free Speech and Cheap Talk. *Journal of Legal Analysis* 46-103.
7. Heymann, Laura A. 2012. The Law of Reputation, and the Interest of the Audience, *B.C. L. Rev.* 1341-1999
8. Koosha Khalvati et al., Modeling other minds: Bayesian inference explains human choices in group decision-making, 5 *Science Advances* (2019)
9. Pornpitakpan, Chanthika. 2004. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence, 34 *Journal of Applied Social Psychology* 243.
10. 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 (11): 651– 657.
11. Lee, S. (2012). A pragmatic analysis of defamation and slanderous remarks. *Ene*, 37(2), 401–416.
12. McNamara, Lawrence. 2007. *Reputation and Defamation*
13. Bar-Gill, Oren and Assaf Hamdani. 2003. Optimal Liability for Libel. *Contributions in Economic Analysis & Policy*. 2(1)
14. Garoupa, Nuno. 1999. Dishonesty and Libel Law: The Economics of the 'Chilling' Effect, *JITE* 284-300
15. Garoupa, Nuno. 1999, The Economics of Political Dishonesty and Defamation, *International Review of Law & Economics*,
16. Dalvi, Manoj and James F. Refalo. 2007. An Economic Analysis of Libel Law. *Eastern Economic Journal*. 74-94