Libel Bullies, Defamation Victims and Litigation Incentives*

David J. Acheson† University of Kent

Ansgar Wohlschlegel‡ Swansea University

Abstract

In the context of defamation law, we analyze a public figure's incentives to bring negative-value defamation suits in order to appear litigious, and how this affects her incentives to do wrong in the first place and the media's incentives to expose this wrongdoing. In equilibrium, the public figure's litigation incentives depend both on her own direct costs and benefits of doing so, and on journalists' costs and benefits from litigation and publication. Furthermore, equilibrium wrongdoing and publication choices depend on an otherwise non-litigious public figure's litigation payoffs. Potential effects of legal reform are briefly discussed.

JEL code: K19, D73

Keywords: Defamation Law, Litigation Costs, Signalling.

*We are grateful to Gerd Mühlheußer, the editor, and two anonymous referees, as well as Eberhard Feess, Murat Mungan and seminar participants at Portsmouth Business School for helpful comments.

†Centre for Journalism, University of Kent, Gillingham Building, University of Kent at Medway, Chatham Maritime, Kent ME4 4AG, United Kingdom. Email: d.acheson@kent.ac.uk

‡School of Management, Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, United Kingdom. Email: a.u.m.wohlschlegel@swansea.ac.uk

1 Introduction

During the 2016 US presidential election campaign, the notoriously litigious Donald Trump declared that he would 'open up' US libel laws to allow public figures to sue media organisations more easily.[1] Trump is not the only high-profile figure to have called for this kind of reform: Supreme Court Justice Clarence Thomas used his concurring opinion in the denial of certiorari in *McKee v Cosby* (2019) to challenge the longstanding first amendment defamation jurisprudence that stems from the Court's seminal decision in *New York Times v Sullivan* (1964), which is the reason that it is uniquely difficult for public figure plaintiffs to succeed with defamation claims in U.S. courts. More broadly, a potential defamation victim's incentives to take a media outlet to court are important to many contemporary political and legal discussions. For example, one of the central concerns raised by the #MeToo movement is that the use of aggressive legal tactics can silence victims of sexual assault or harassment, and prevent the exposure of even serial abusers in the media.[2]

Society would like to encourage genuine victims to sue for defamation in order to deter media from publishing false and libelous claims. On the other hand, if defamation suits are too attractive for plaintiffs, wrongdoers exposed by true media stories may sue, thus deterring even truthful publications. Ironically, towards the end of the aforementioned election campaign, Trump was involved in a slightly bizarre story when the American Bar Association refused to publish a report concluding that Trump was a 'libel bully' because of concern about the possibility that Trump would sue for libel.[3] More recently, Trump's 2020 campaign has been criticized for using 'utterly frivolous'[4] defamation suits, filed against outlets such as The New York Times, CNN, and The Washington Post, 'to attack and intimidate America's journalists'.[5] This kind of deterrence of legitimate publications, commonly referred to as the 'chilling effect' (Schauer, 1978), prevents journalists from fulfilling their function of holding public figures accountable for their actions, so that wrongdoing goes unexposed and, therefore, undeterred.

However, incentives to sue for libel and how they are interrelated with wrongdoing and the veracity of media stories in equilibrium are far from obvious. In particular, if journalists can only observe noisy evidence of a public figure's wrongdoing, which cannot be perfectly verified or falsified even in court, this lack of observability makes it impossible for defamation law to implement the desirable outcome that stories are published if and only if they are true. The uncertainty as to the outcome of litigation means that the anticipated likelihood and potential

[1] Hadas Gold, 'Donald Trump: We're Going to "Open Up" Libel Laws' (Politico, February 26 2016).

[2] Kara Fox and Antoine Crouin, 'Men are suing women who accused them of harassment. Will it stop others from speaking out?' (CNN, June 5 2019).

[3] Mark Joseph Stern, 'American Bar Association Produces Report Calling Trump a Libel Bully, Censors It Because He's a Libel Bully' (Slate, October 25 2016).

[4] Matthew Rozsa, 'Laurence Tribe: Trump Campaign's Lawsuit against New York Times "Designed to Chill the Free Press"' (Salon, 27 February 2020; quoting Laurence Tribe).

[5] Joshua A Geltzer and Neal K Katyal, 'The True Danger of the Trump Campaign's Defamation Lawsuits' (The Atlantic, 11 March 2020). See also: Theodore J Boutrous Jr, 'Why Trump's Frivolous Libel Lawsuit against the New York Times is Dangerous' (The Washington Post, 29 February 2020).

cost of a defamation suit assume greater importance in the journalist's publication decision. The decision whether to publish a story based on this noisy evidence will largely depend on what each player anticipates the other player to do in equilibrium: that is, the journalist will publish if he believes that the public figure will not sue in equilibrium, rather than if he believes that the story is true.

Furthermore, a public figure's incentives to sue for libel does not only depend on the expected costs and direct benefits of the lawsuit in terms of potential damages or the value of the original allegations being withdrawn or their continuing publication stopped. There may be other journalists in the future who observe evidence of some other wrongdoing, and the public figure's reaction to previous allegations will give them a hint of how likely she is to sue in subsequent cases. As we have just argued that journalists' publication decisions are heavily influenced by their expectations of the likelihood that they will be sued, a public figure may anticipate this and sue for libel even if the costs of a lawsuit exceed the direct benefits, just to discourage future publications by other journalists.

The aim of this paper is to present an economic model that accommodates these intuitive ideas and to derive avenues for legal reform to address these issues. In particular, we are interested in the relationship between public figures' incentives to sue for defamation and journalists' incentives to self-censor critical reporting because of their fear of being sued. We consider a public figure that may do wrong in each of two periods. In each period, there is a journalist who may find evidence of the public figure's wrongdoing. However, this evidence is noisy: evidence is produced with certainty if the public figure has done wrong, but also with some probability if she has not. If the journalist publishes a story based on this evidence claiming that wrongdoing has occurred, the public figure may sue for defamation. Both litigants' net payoffs from this lawsuit depend on whether the story is true. Furthermore, the public figure may be a litigious type, who always benefits from litigation, or a non-litigious type who always incurs a cost from litigation. This difference may be due to differences in, for instance, the extent to which each type would benefit from vindicating her reputation should the court find in her favor; their access to high-quality legal advice; or their personal disutility from being involved in lawsuits. Whatever the source of the difference between the types, we assume that each public figure's type is her private information.

As our aim is to analyze the situation where the chilling effect can exist, i.e., where the threat of litigation is significant enough to be capable of preventing the journalist from publishing true allegations, we assume that journalists' net litigation costs exceed publication benefits even if the story is true. Anecdotal evidence from both sides of the Atlantic suggests that this is a plausible assumption: In the UK, litigation costs affect the publication decisions even of large media companies, such as the BBC or the Guardian (Weaver, Kenyon, Partlett, and Walker (2004), pp. 1312-1313), and smaller publishers may be forced to retract even true statements of public interest (Lopez (2012)). In the US, the point that the chilling effect of defamation law is caused by the cost of litigation itself, and not just the publisher's risk of losing in court, has

been recognized for several decades.[6]

In the main version of the model, we will assume that the second-period journalist can observe first-period publication and litigation decisions and use this to make inferences about the public figure's type, which inform his decision whether to publish allegations against her. This, in turn, will create incentives for the low-litigation benefit type to imitate the litigious type in the first period. Our analysis will also deliver results for a version of the model in which the second-period journalist cannot observe the first-period publication and litigation decisions. Therefore, we will be able to compare the results of both versions of the model to learn more about the specific impact of reputational concerns on equilibrium.

Our main results are that, *ceteris paribus*, litigation incentives are higher when public figures' litigation histories are known to journalists, as plaintiffs bring negative-value suits in order to appear litigious. Furthermore, the incentive for an otherwise non-litigious type of public figure to imitate libel bullies introduces an impact of her litigation costs on equilibrium that is not present in a one-period model. We will demonstrate in Section 5 the relevance of this effect for legal reform when, for instance, improving the accuracy of defamation courts.

A distinctive feature of defamation cases is that the potential plaintiff is identifiable by the potential defendant even before the allegedly tortious conduct (in this case, the publication of defamatory allegations) takes place. Therefore, by building up a reputation for litigiousness, the plaintiff can prevent that conduct altogether.[7] By contrast, in most other torts, lawsuits are based on actions taken in one-off encounters, so that a potential plaintiff's litigiousness is not being taken into account by the potential defendant. For instance, when deciding how careful to be when driving a car, a driver will not consider potential differences in other drivers' litigiousness. In those other cases, a reputation for being litigious may still be valuable, but only once the disputed action has taken place. Previous papers have discussed bringing nuisance suits in order to extract a high settlement (Farmer and Pecorino (1998)) or rejecting a settlement in order to deter future nuisance suits (Miceli (1993)).

Note that the incentive structure of our model of potentially wrongdoing public figures who may or may not sue journalists for defamation bears some resemblance with that in a game of whistleblower protection, whereby potentially wrongdoing firms may dismiss or demote employees who blow the whistle. Whistleblowers may benefit via formal rewards (see Givati, 2016) or via personal motives such as conscience cleansing or punishing the wrongdoer (see Heyes and Kapur, 2008). On the other hand, whistleblowing may trigger an investigation which will inflict harm on the wrongdoer down the line (Heyes and Kapur, 2008). Whistleblower

[6]"...the relevant question is not whether a story is libelous, but whether the subject is likely to sue" (Anderson (1975), p. 425); "The chilling effect comes from the threat of lawsuits which may not be won [by the plaintiff] but which will nevertheless impose an enormous (and, generally under our present system, unrecoverable) cost" (Boies (1994), p. 1210); "...the threat of successful libel action is not necessarily what influences publications. Rather, it is the threat of any litigation." (Bergelson (2018), p. 229)

[7]Other examples for such a setting are patent litigation, as shown by Hovenkamp (2013), and the treatment of whistleblowers in employment relations, as discussed further below.

protection law may impose costs on an employer who punishes a whistleblower in some of the aforementioned ways. An exposed employer may still choose to incur the cost of punishing the whistleblower in order to discourage future whistleblowing. At the same time, asserting whistleblower protection is costly for an employee and particularly difficult if punishment is informal such as mobbing by co-workers, as opposed to formal action taken by an employer such as dismissal or demotion. In this setting, our model's contribution is to stress how an exposed employer's incentives to punish a whistleblower depends on future whistleblowers' costs and benefits of blowing the whistle and, potentially, litigating over their protection. To the best of our knowledge, this effect has not yet been identified in the existing literature on whistleblower protection.

The most closely related paper is Garoupa (1999a), who also analyzes the impact of libel law on wrongdoing and publication incentives, but assumes that a public figure automatically sues for defamation if a story is published, and does not consider the public figure to act repeatedly. Other economic models of libel law discuss the media's incentives to invest in the accuracy of the evidence that their story is based on (Bar-Gill and Hamdani (2003) and Dalvi and Refalo (2008)), anonymous sources' incentives to leak stories to journalists under different burdens of proof for the journalist (Baum, Feess, and Wohlschlegel (2009)), and voters' reactions to corrupt politicians exposed by the media (Gratton (2015)). None of these papers studies litigation incentives and their relation to wrongdoing and publication incentives, or the impact of the observability of a public figure's litigation history on these issues.[8]

In general, an aim of libel law from an economic point of view is to align a defendant's interests with social welfare, which it has in common with tort law. Indeed, many of the existing economic models on libel law are driven by the same logic as models on tort law more generally. For instance, the main driving force in Bar-Gill and Hamdani (2003) is the trade-off between promoting care and activity levels which can be readily applied to other areas of tort law, as in Feess, Muehlheusser, and Wohlschlegel (2009). However, we focus on the unique feature of defamation cases that the media's aim of delivering stories of public interest means that many stories specifically target public figures who are in the spotlight for a longer period of time. Our paper, thus, deviates from the notion of anonymous encounters usually assumed in accident models.[9]

Our model builds on two concepts from the legal literature.[10] Firstly, we have already mentioned the chilling effect theory, according to which publishers' fear of defamation suits

[8]Garoupa (1999b) does analyze litigation incentives, but in a model in which the public figure's wrongdoing is perfectly observable to the journalist.

[9]Our model may be superficially perceived as similar to tort models where there are multiple potential defendants in the same case who act sequentially (e.g. Kim and Song (2007) or Kim and Lee (2019)). However, in contrast to our model, injurers in those cases typically impose harm on the victim in an interdependent way, and once the case goes to court, all potential injurers have made their choice of care level. Therefore, the results from that literature are not readily applicable to our model.

[10]See Acheson and Wohlschlegel (2018a) for a more detailed account of these concepts and their relation to the economic literature, including this paper.

may deter them from publishing not only false stories, but also true stories. Chilling effect reasoning has influenced important judgments in the highest courts of many countries.[11] The idea that defamation law has some influence on the media's publishing decisions, and that laws unfavourable to journalists may have a chilling effect on their speech, is well-established in the academic legal literature: see, for instance, Barendt, Lustgarten, Norrie, and Stephenson (1997), Anderson (1975), or Weaver, Kenyon, Partlett, and Walker (2006). In particular, the central importance of the high cost of defending libel actions, on which our model is partly based, is well-recognized: see Schauer (1978), or Mullis and Scott (2009). The legal literature mainly focuses on the chilling effect of defence costs on the output of institutional media, as for instance in Dent and Kenyon (2004), but Townend (2014) suggests that similar effects operate on some small online publishers as well.

Secondly, our model assumes that a public figure's perceived litigiousness influences journalists' decisions whether to publish allegations of wrondoing against her. While there is little direct evidence to support this claim, indirect and anecdotal evidence suggest that this is in fact the case, such as Barendt, Lustgarten, Norrie, and Stephenson (1997), Weaver (2012) for UK publishers or Kenyon and Marjoribanks (2008) for Australian publishers. One of the most commonly mentioned high-profile figures perceived as abusing libel laws to stifle criticism of their activities is Robert Maxwell, the former proprietor of the Mirror Group press company in the UK. As Hooper (2000) notes, although Maxwell's success rate in court was poor, his threats of litigation were successful in suppressing critical media coverage. The result was that Maxwell's serious financial crimes were not exposed by the British press until after his death.[12] Similar tactics have helped to hide the wrongdoing of public figures such as Lance Armstrong and Jimmy Savile (Townend (2019)), and to stifle criticism of companies like McDonald's (Vick and Macpherson (1995)) and nonprofit organizations like the Church of Scientology (Boies (1994), p. 1209).

The remainder of the paper is organized as follows: We start by setting out the model assumptions. We then analyze equilibrium play in the second period for some exogenously given beliefs that the second-period journalist has about the public figure's type. The results of this section can be readily applied to a version of the model in which later journalists do not observe a public figure's litigation history. In section 4, we analyze the full game. Section 5 will use these results to discuss potential avenues for legal reform and Section 6 concludes.

[11] For instance, *New York Times Co v Sullivan* (1964) in the US, *Derbyshire County Council v Times Newspapers Ltd* (1993) in the UK, *Theophanus v Herald & Weekly Times Ltd* (1994) in Australia, *Grant v Torstar Corp* (2009) in Canada, or *Independent News & Media v Ireland* (2006) in the supranational European Court of Human Rights.

[12] See Vick and Macpherson (1995).

2 The Model

Players and Timing. The main purpose of the model is to analyze a journalist's optimal reaction to evidence of wrongdoing and how the alleged wrongdoer has previously been observed to behave in a similar situation, and a potential wrongdoer's optimal choice anticipating the journalist's strategy. To this end, we consider a two-period game[13] with three players: A public figure P chooses, in each period, whether to do some wrong. In each period, there is a different journalist, denoted J_1 in period 1 and J_2 in period 2, who receives a noisy signal ('evidence') of whether P has done wrong and may, depending on this signal, publish a story to allege this wrongdoing. In reality, even if they turn out to be untrue, media stories are not usually just made up out of thin air but rather based on some evidence. We acknowledge this in our model by assuming that journalist J_t cannot publish a story unless some evidence (the noisy signal) has emerged in period t indicating that P might have done wrong.

If a news story is published, P suffers a loss r of reputation, and P may decide whether to sue the journalist for libel. There are two types of public figures, where type H's payoff from a libel lawsuit is larger than type L's independent of whether the allegation of wrongdoing is true. Journalists cannot observe the public figure's type but only the prior probability for each type.[14] In reality, many drivers of a public figure's cost and benefit from litigation will be publicly observable. However, all we need for our model is that some of them are not.

A useful example for this may be the extent to which a public figure benefits from the retraction of a published story. While some celebrities' popularity does not depend much on whether reporting on them is positive or negative, as long as they are in the news, other public figures, such as politicians or entrepreneurs, heavily depend on a public image of integrity and trust, which might be damaged by an allegation of wrongdoing. Such an allegation may have severe consequences for this latter type of public figure's future income or career unless it is publicly retracted. However, the extent of this harm depends on that public figure's future business or career plans, which a journalist who contemplates publishing a story on that person would not know about. Other determinants of the public figure's expected cost and benefits from litigation that are privately known to her include her personality (such as her vindictiveness) or her opportunity cost of the time needed to engage with the lawsuit.

In summary, the timeline is as follows:

0 Nature determines potential wrongdoer P's type $i \in \{L, H\}$, which is private information to P. The ex-ante probability that $i = H$ is g.

[13] It will turn out that the main distinction between both periods is whether the game ends after that period. Therefore, introducing more periods would not change the model fundamentally. All periods except the final one would be qualitatively similar to the first period in our model.

[14] We label the players in our model 'journalists' and 'public figure', respectively, as journalists are typically better at detecting news stories than the general public, and because it is typically public figures whose reputation suffers from being exposed as wrongdoers and whose wrongdoing society needs to be accurately informed about. However, the model is equally applicable to any such players with these characteristics.

Period $t \in \{1, 2\}$:

t(i) P decides on whether to do wrong $w_t \in \{0, 1\}$ in Period t. We denote the case of wrongdoing by $w_t = 1$.

t(ii) Noisy signal $s_t \in \{0, 1\}$ sent to Period-t journalist J_t on whether P has done wrong in step t(i). If P has done wrong in step t(i), then $s_t = 1$ with certainty, whereas if P has not done wrong, $s_t = 1$ with probability σ and $s_t = 0$ with probability $1 - \sigma$.[15]

t(iii) If $s_t = 1$, J_t decides whether to publish news story $n_t \in \{0, 1\}$, where $n_t = 1$ denotes publication of a news story in period t.

t(iv) If J_t has published, P decides whether to sue J_t for libel, $\gamma_t \in \{0, 1\}$, where $\gamma_t = 1$ means that P sues J_t in period t.

t(v) Payoffs are realized depending on the players' actions.

In Period 0, the potential wrongdoer's type is determined for the entire game. Each of the subsequent Periods $t \in \{1, 2\}$ are subdivided in five steps $t(i) - t(v)$. We assume that all publication and litigation decisions are publicly observable, whereas the wrongdoing decisions and the nature move in stage 0 are only known to the public figure.

Payoffs. In each period $t \in \{1, 2\}$, a type-i potential wrongdoer's expected payoff in period t is defined as

$$\Pi_t^i = w_t b - n_t r + \gamma_t \ell_{w_t}^i. \tag{1}$$

If P does wrong ($w_t = 1$), she acquires the benefit from wrongdoing b. However, if journalist J_t publishes a news story ($n_t = 1$), P suffers reputational harm r. If P sues J_t for libel, her expected net benefit from this lawsuit is, depending on whether she had actually done wrong, $\ell_{w_t}^i$, which includes the expected benefit from the story being withdrawn, expected damages and expected legal expenses.[16]

Similarly, journalist J_t's expected payoff is

$$\Pi_t^J = n_t p + \gamma_t \ell_{w_t}^J. \tag{2}$$

If journalist J_t publishes a news story ($n_t = 1$), he obtains benefit p. However, if the public figure sues him subsequently ($\gamma_t = 1$), J_t's expected net benefit $\ell_{w_t}^J$ from the litigation game,

[15] While the possibility that the signal s_t produces false positives with some probability is necessary for type L to imitate type H, our results do not qualitatively depend on whether false negatives are possible. For the sake of parsimony, we, therefore, rule this possibility out by assumption.

[16] We follow the literature on the economics of defamation law by assuming that the general public is passive, so that r and ℓ_k^i are exogenous parameters. However, in reality the general public might observe litigation decisions and outcomes and react rationally to these observations, which would make these payoffs endogenous. We will briefly discuss this assumption in Section 6.

which includes the expected cost of withdrawing the story, expected damages and expected legal expenses, is added.

We confine our analysis to the parameter range where it differs from a single-period model by making the following assumptions:[17]

Assumption 1 *(i)* $\ell_1^L \leq \ell_0^L < 0 < \ell_1^H \leq \ell_0^H < r$.

(ii) $b < (1-\sigma)r$.

(iii) $p < -\ell_1^J \leq -\ell_0^J$.

For a public figure who has indeed done wrong, a lawsuit cannot be more profitable than if she had not done wrong. Furthermore, we model type H as a libel bully by assuming that her net benefit from a libel lawsuit is always positive,[18] whereas a lawsuit is assumed to always be costly for type L. Last, we assume that even a non-wrongdoing type H's reputation cannot be fully restored in court ($\ell_0^H < r$). An advantage of this last assumption is that we avoid the counter-intuitive situation in which type H wants to imitate type L in order to tempt J_t into publishing a news story just to be able to sue him.

The second part of the assumption reflects the intuitive notion that wrongdoing does not pay off for type L if she anticipates that J_t publishes with certainty. In other words, this condition assumes that well-functioning media are an effective deterrent against wrongdoing. In the last part of the assumption, we assume that being taken to court harms the journalist in expectation, no matter whether the story is true or false, and that, if the journalist is certain he will be sued, the expected costs of litigation will always outweigh the benefits from publication.

Strategies and Beliefs. Allowing for mixed strategies, let ω_t^i, $t = 1, 2$, $i \in \{L, H\}$ denote the probability with which type i does wrong in period t, η_t, $t = 1, 2$ the probability with which a journalist in period t publishes a story upon receiving wrongdoing signal, λ_w^i the probability with which type i sues for defamation after having done wrong ($w = 1$) or not having done wrong ($w = 0$), and μ_t the probability with which a period-t journalist believes the alleged wrongdoer to be of type H.[19] Note that, while the first-period journalist J_1 has only one occasion to update his beliefs (from g to μ_1 upon observing $s_1 = 1$), there are multiple occasions at which the second-period journalist J_2 does so: Observing whether or not J_1 publishes, whether or not the public figure sues J_1 after he has published, and observing $s_2 = 1$ all reveal information about the public figure's type. We must, therefore, be more specific about which point in time beliefs μ_2 refer to: We define μ_2 as J_2's beliefs just before observing s_2. Note that $\mu_2 = \mu_2(n_1, \gamma_1)$, is a

[17] We will briefly discuss equilibrium when relaxing these assumptions at the end of Section 4.

[18] The case $\ell_1^H < 0 < \ell_0^H$, where type H benefits from litigation only if the allegation is false, involves some interesting strategic trade-offs for type H and is probably more in line with real-world libel bullies' behaviour, who do not sue over all negative coverage all the time, but yields similar results as our model in terms of incentives for type L to imitate type H. As the paper's focus is on type L's decision making, we decided to confine the analysis to the case where H always wants to file suit.

[19] Note that there is a decision for J_t to make only if $s_t = 1$, which is why μ_t is only defined in this case.

function of the entire signal and litigation history up to that point in time.[20] This inference that journalists in the later period make from observing a potential wrongdoer's earlier decisions is the driving force for the result that it might pay off for a public figure to file a negative-value lawsuit just to appear litigious.

3 Final Period

Our analysis starts with the choices in period 2 that can occur in a Perfect Bayesian equilibrium depending on μ_2, J_2's beliefs based on observing all publicly observable outcomes in period 1, but before observing the period-2 signal s_2. Note this is equivalent to a one-period version of the model when setting $\mu_2 = g$.

Let us start by analyzing P's decision to sue for defamation in period 2. Recall that this is an option only if the period-2 journalist J_2 has published a story alleging P's wrongdoing, which in turn requires that the period-2 signal $s_2 = 1$. In this case, P prefers suing for defamation if and only if $\ell^i_{w_2} > 0$.

Due to our assumption $\ell^L_1 < 0 < \ell^H_0$, P prefers suing for defamation if and only if she is type H. This condition already represents the main driving force of this paper's argument: Absent any signalling considerations, the high-type public figure will sue for defamation, whereas the low type will not. Hence, if the journalist knew the public figure's type, for instance, by inference from the public figure's earlier decisions, he would publish a story about her wrongdoing if and only if he believes that she is the low type. However, as a type L public figure who knows that she has been exposed as such has no incentive to do wrong, the libel system seems to promote a publication strategy that goes contrary to the journalist's belief of whether the story is true. The following analysis of the model will further illustrate this argument.

Consider now J_2's decision to publish upon observing $s_2 = 1$. J_2's payoff when not publishing is zero. His expected payoff from publishing depends on his initial beliefs μ_2 before observing the signal and on how he updates these beliefs using both types of public figure's equilibrium strategies ω^H_2 and ω^L_2: He will publish if and only if

$$p + Prob(i = H \wedge w^H_2 = 1 \mid s_2 = 1)\ell^J_1 + Prob(i = H \wedge w^H_2 = 0 \mid s_2 = 1)\ell^J_0 \geq 0. \tag{3}$$

With Bayes' rule, this is equivalent to

$$p + \frac{\mu_2\omega^H_2\ell^J_1 + \mu_2(1-\omega^H_2)\sigma\ell^J_0}{\mu_2[\omega^H_2 + (1-\omega^H_2)\sigma] + (1-\mu_2)[\omega^L_2 + (1-\omega^L_2)\sigma]} \geq 0. \tag{4}$$

The fraction on the left-hand side of (4) weights the journalist's expected payoffs from the libel lawsuit, depending on whether the story is true, with the joint probabilities of the story being true or false, P's type being H and evidence being observed. (4) implies that J_2 is more likely to publish if his initial beliefs μ_2 are low, as this makes P less likely to be type H and, thus,

[20] n_1 must be an argument of $\mu_2(.)$, because the alleged wrongdoer can sue J_1 for libel only if a news story has been published. Hence, $\gamma_1 = 0$ can serve as a signal for the alleged wrongdoer being type L only if $n_1 = 1$.

to sue for libel; if ω_2^H is high, as this makes the story more likely to be true, in which case J_2's expected payoff from the lawsuit is higher; or if ω_2^L is high, as this makes P less likely to be type H conditional on observing the signal $s_2 = 1$.

Next, consider P's decision whether to do wrong in period 2. Each type's expected payoff from wrongdoing will depend on the probability η_2 of J_2 publishing upon observing $s_2 = 1$. In particular, (1) implies that the public figure prefers doing wrong if and only if

$$\text{Type } H: \qquad b - \eta_2 \left(r - \ell_1^H\right) \geq -\sigma\eta_2 \left(r - \ell_0^H\right) \tag{5}$$

$$\text{Type } L: \qquad b - \eta_2 r \geq -\sigma\eta_2 r. \tag{6}$$

Each type needs to take into account that not doing wrong does not necessarily prevent J_2 from publishing: With probability σ, signal $s_2 = 1$ will be observed although P has not done wrong, in which case J_2 publishes with probability η_2.

Combining the conditions (4), (5) and (6) shows how the wrongdoing and publication decisions are interrelated: If J_2 always publishes ($\eta_2 = 1$), type L never wants to do wrong. However, this increases J_2's ex-post beliefs of the public figure being type H, so that J_2 prefers not publishing if μ_2 is sufficiently high. On the other hand, if J_2 never publishes, both types of public figure always want to do wrong, which reduces J_2's ex-post beliefs of the public figure being type H, so that J_2's publishing may be restored if μ_2 is sufficiently low. Hence, we expect intuitively the equilibrium to be that J_2 always (never) publishes and type L never (always) does wrong if μ_2 is low (high).

To analyze equilibrium more formally, let us focus, for the moment, on the special case where

$$0 < \sigma\ell_0^H + (1-\sigma)r - b \leq \ell_1^H, \tag{7}$$

which implies that type H prefers doing wrong even if it is known that J_2 always publishes ($\eta_2 = 1$). Substituting for $\omega_2^H = 1$ in (4) yields

$$p + \frac{\mu_2 \ell_1^J}{\mu_2 + (1-\mu_2)[\omega_2^L + (1-\omega_2^L)\sigma]} \geq 0. \tag{8}$$

Furthermore, (6) implies that type L prefers doing wrong if and only if $\eta_2 \leq \frac{b}{r(1-\sigma)}$. As, by assumption, $b \leq (1-\sigma)r$, L always (never) does wrong if J_2 never (always) publishes. Hence, whether either case is an equilibrium depends on J_2's publication incentives.

If L always does wrong, (8) implies that J_2 prefers not publishing if $\mu_2 > \frac{p}{-\ell_1^J}$, which is, therefore, the necessary and sufficient condition for a pure-strategy equilibrium in which L always does wrong and J_2 never publishes. Similarly, substituting for $\omega_2^L = 0$ in (8) yields $\mu_2 < \frac{\sigma p}{-\ell_1^J - (1-\sigma)p}$ as the necessary and sufficient condition for a pure-strategy equilibrium in which L never does wrong and J_2 always publishes.

A mixed strategy equilibrium in which a type L public figure randomizes between doing wrong and not doing wrong requires that (6) holds with equality for some $\eta_2 \in [0,1]$, i.e. that it is satisfied for $\eta_2 = 0$ (which is true by assumption) but violated for $\eta_2 = 1$. For (6) to hold with

equality, $\eta_2 \in (0,1)$ typically is required, which in turn requires that J_2 is indifferent between publishing and not publishing. This is possible only if (4) is satisfied for $\omega_2^L = 1$ but violated for $\omega_2^L = 0$. Summing up, this is an equilibrium if $\frac{\sigma p}{-\ell_1^J-(1-\sigma)p} \leq \mu_2 \leq \frac{p}{-\ell_1^J}$. On the edges of this range, J_2 being indifferent between publishing and not publishing is supported by a pure strategy of type L, i.e. $\omega_2^L = 0$ if $\mu_2 = \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}$, and $\omega_2^L = 1$ if $\mu_2 = \frac{p}{-\ell_1^J}$.

Summing up, the following second-period strategies are consistent with a given initial second-period belief μ_2:

(i) If $\mu_2 < \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}$, type L does not do wrong and J_2 publishes. Type L's expected second-period payoff is $\pi_2^L(\mu_2) = -\sigma r$.

(ii) If $\mu_2 = \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}$, type L does not do wrong, and J_2 publishes with probability $\eta_2 \in \left[\frac{b}{r(1-\sigma)}, 1\right]$. Type L's expected second-period payoff is $\pi_2^L(\mu_2) = -\sigma\overline{\eta}_2 r$.

(iii) If $\frac{\sigma p}{-\ell_1^J-(1-\sigma)p} < \mu_2 < \frac{p}{-\ell_1^J}$, type L randomizes between doing wrong and not doing wrong, and J_2 publishes with probability $\eta_2 = \frac{b}{r(1-\sigma)}$. Type L's expected second-period payoff is $\pi_2^L(\mu_2) = -\frac{b\sigma}{1-\sigma}$.

(iv) If $\mu_2 = \frac{p}{-\ell_1^J}$, type L does wrong, and J_2 publishes with probability $\eta_2 \in \left[0, \frac{b}{r(1-\sigma)}\right]$. Type L's expected second-period payoff is $\pi_2^L(\mu_2) = b - \underline{\eta}_2 r$.

(v) If $\mu_2 > \frac{p}{-\ell_1^J}$, type L does wrong, and J_2 does not publish. Type L's expected second-period payoff is $\pi_2^L(\mu_2) = b$.

There are two threshold values for μ_2 below the lower (above the upper) of which L never (always) does wrong and J_2 never (always) publishes. Furthermore, in cases (ii) and (iv), a continuum of mixed strategies for J_2 support equilibrium, which makes type L's equilibrium payoff continuous in μ_2 in the sense that for any possible payoff $\widetilde{\pi} \in [-\sigma r, b]$ we can find some μ_2 such that type L's expected second-period equilibrium payoff is $\pi_2^L(\mu_2) = \widetilde{\pi}$. This will play an important role in Section 4 when constructing a mixed strategy equilibrium where type L randomizes between suing and not suing in the first period. The following lemma shows that these qualitative characteristics of this equilibrium hold more generally whenever our assumption $\ell_1^H > 0$ is satisfied:

Lemma 1 *The equilibrium strategies in the second period, depending on μ_2, are as follows: Type H always and type L never sues for defamation. There exist $\widetilde{\mu}^1$ and $\widetilde{\mu}^2$ with $\widetilde{\mu}^1 < \widetilde{\mu}^2$ such that type L never does wrong if $\mu_2 \leq \widetilde{\mu}^1$, always does wrong if $\mu_2 \geq \widetilde{\mu}^2$ and randomizes between doing and not doing wrong if $\widetilde{\mu}^1 < \mu_2 < \widetilde{\mu}^2$.*

(i) If $\ell_1^H \geq \sigma\ell_0^H + (1-\sigma)r - b$, type H always does wrong. J_2 always publishes if $\mu_2 < \widetilde{\mu}^1$, never publishes if $\mu_2 > \widetilde{\mu}^2$ and randomizes between publishing and not publishing if $\widetilde{\mu}^1 \leq \mu_2 \leq \widetilde{\mu}^2$.

(ii) If $0 < \ell_1^H < \sigma\ell_0^H + (1-\sigma)r - b$, there exists $\widetilde{\mu}^0 \leq \widetilde{\mu}^1$ such that type H never does wrong if $\mu_2 \leq \widetilde{\mu}^0$, always does wrong if $\mu_2 \geq \widetilde{\mu}^1$ and randomizes between doing and not doing

wrong if $\widetilde{\mu}^0 \leq \mu_2 \leq \widetilde{\mu}^1$. J_2 always publishes if $\mu_2 < \widetilde{\mu}^0$, never publishes if $\mu_2 > \widetilde{\mu}^2$ and randomizes between publishing and not publishing if $\widetilde{\mu}^0 \leq \mu_2 \leq \widetilde{\mu}^2$.

In both of these cases, $\widetilde{\mu}^i$ are nondecreasing in p, σ, ℓ_0^J and ℓ_1^J and do not depend on any other parameters, and there exists, for every $\widetilde{\pi} \in \{-\sigma r, b\}$, a second-period belief μ_2 such that type L's equilibrium second-period profit is $\pi_2(\mu_2) = \widetilde{\pi}$. Furthermore, for every $\mu_2 = \widetilde{\mu}^i$, there is a continuum of equilibrium publishing probabilities for J_2.

Proof. See the Appendix. ■

Lemma 1 confirms our intuition, explained in the Introduction, that a journalist is more inclined to publish if he thinks the public figure is less likely to be the litigious type H. In other words, his publication decision is driven by his expectation of the public figure's litigation decision rather than the veracity of the story: To the contrary, the range where J_2 publishes with certainty is even larger if the journalist's evidence is less accurate (σ high), as this means that the public figure is more likely to be type L if that type does not do wrong in equilibrium. Furthermore, wrongdoing is less and publication more likely if the journalist's benefit p from publishing is higher and the costs from being dragged into a lawsuit $-\ell_1^J$ are lower.

On the one hand, Lemma 1 is a necessary exercise to prepare the analysis of the full game. On the other hand, however, recall that this result is important in its own right, as it can be interpreted as the equilibrium of a one-period version of our game, or a version in which public figures are not recognizable by later journalists, when substituting for $\mu_2 = g$.

4 First Period

Decision to Sue When deciding whether to sue in the first period after doing wrong ($k = 1$) or not doing wrong ($k = 0$), type L takes into account what J_2 will believe after observing the first-period play. As $\ell_1^H \geq 0$, type H always sues in either period, so that J_2's rational beliefs about whether the public figure is type H are

$$\mu_2 = \mu_2^S := \frac{g(\omega_1^H + (1-\omega_1^H)\sigma)}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)(\omega_1^L\lambda_1^L + (1-\omega_1^L)\sigma\lambda_0^L)} \tag{9}$$

if a story was published in the first period and the first-period journalist was sued, and $\mu_2 = \mu_2^N := 0$ if a story was published but went unchallenged by the public figure. The denominator of (9) consists of the probability g that the public figure is type H, who always sues, and J_1 observes a signal about her, and the probability that the public figure is type L and J_1 observes a signal, taking into account that type L's decision to sue for defamation may be different after having done wrong (λ_1^L) rather than after having not (λ_0^L).

On the other hand, not suing in the first period immediately identifies the public figure as type L. In this case, J_2 will publish whenever observing the signal. Even if L anticipates this and does not do wrong in the second period, evidence may still be produced due to the noisiness of the signal. In this case, J_2 will still publish, although he can infer from the fact that the

public figure is type L that her optimal decision was not to do wrong. Again, J_2's incentives to publish are inversely related to his beliefs that the story is true.

To sum up, J_2 believes that the public figure is more likely to be type H after suing in the first period than after not suing ($\mu_2^S > \mu_2^N$). According to Lemma 1, higher beliefs μ_2 imply that the public figure is more likely to do wrong and J_2 less likely to publish. Hence, the public figure's payoff is also higher if J_2 believes that she is type H with higher probability, so that type L benefits from imitating H by suing for defamation in the first period.

In other words, type L trades off the litigation cost $-\ell_k^L$ with a potentially higher second-period payoff $\pi_2^L(\mu_2^S)$ after suing rather than $\pi_2^L(\mu_2^N)$ after not suing: If type L's litigation cost $-\ell_k^L$ is sufficiently large, it is optimal for her not to sue ($\lambda_1^L = \lambda_0^L = 0$), so that J_2's rational beliefs after observing a lawsuit in the first period are $\mu_2^S = 1$. As the public figure's second-period payoff is increasing in J_2's beliefs, these beliefs $\mu_2^S = 1$ create the highest possible incentives for suing for defamation in the first period. If type L prefers not to sue even when anticipating J_2 to have these beliefs, and after not doing wrong in the first period, i.e., if $-\ell_0^L \geq \pi_2^L(1) - \pi_2^L(0)$, then type L never sues for defamation in the first period in equilibrium.

Recalling from Lemma 1 that, in equilibrium, L never (always) does wrong in the second period and J_2 always (never) publishes if $\mu_2 = 0$ ($\mu_2 = 1$), type L's second-period payoffs are $\pi_2^L(1) = b$ and $\pi_2^L(0) = -\sigma r$. We have, thus, established the following proposition:

Proposition 1 *In equilibrium, the type L public figure never sues a publishing first-period journalist J_1 for defamation if and only if $-\ell_0^L \geq b + \sigma r$.*

This proposition is important as it establishes that for sufficiently high litigation costs there is no incentive for type L to imitate type H, so that there is no difference between this model and a one-period model. On the other hand, for sufficiently small but positive litigation costs, type L will imitate type H by suing with some probability after being accused of wrongdoing by the media. Note, however, that this result crucially hinges on the presence of a true libel bully: If type H's expected litigation payoff ℓ_1^H falls below zero, her and type L's incentive to sue for defamation decrease, and if even ℓ_0^H becomes negative, there is no way for L to make J_2 believe she is type H by suing for defamation in the first period.

Furthermore, whether such a lawsuit actually takes place depends on whether the media make any wrongdoing allegations against the public figure in the first place, which, in turn, depends on the public figure's anticipated litigation strategy. For instance, if it is known that type L always sues in equilibrium, then J_1 would anticipate this and never publish in the first place. Whether the first-period journalist publishes in equilibrium is important as this is the only deterrent against the public figure's wrongdoing. For instance, if J_1 never publishes, both types of public figure always do wrong in the first period. This is in line with the notion of the 'chilling effect' formulated in the legal literature, so that an important goal of the economic analysis of defamation law is to understand how the media's incentives to publish can be retained. In the following, we will, therefore, discuss the impact of our model parameters on wrongdoing in equilibrium.

Wrongdoing Decision The situation that we have just described, that type L always sues, which induces J_1 to refrain from publishing and both types of public figure to do wrong, need not be an equilibrium. In general, J_2's beliefs after observing no publication at all in the first period are

$$\mu_2 = \mu_2^{NP} := \frac{g[1-\eta_1(\omega_1^H + (1-\omega_1^H)\sigma)]}{g[1-\eta_1(\omega_1^H + (1-\omega_1^H)\sigma)] + (1-g)[1-\eta_1(\omega_1^L + (1-\omega_1^L)\sigma)]} \tag{10}$$

where η_1 is the probability that J_1 publishes after observing a signal: No publication is made when no signal is observed or J_1 decides not to publish after observing a signal. When both types of public figure always do wrong ($\omega_1^H = \omega_1^L = 1$), they are indistinguishable for the second-period journalist, so that equilibrium beliefs will be $\mu_2^{NP} = g$. However, if g is sufficiently small, type L's second-period payoff $\pi_2^L(\mu^{NP})$ will be no better than the second-period payoff $\pi_2^L(0)$ in the case where type L deviates by not suing and is identified as type L. In such a parameter constellation, J_1 will anticipate that type L would not have any benefit to justify the costs of a defamation lawsuit, and, as g is low, that the public figure is unlikely to be type H, thus restoring publication incentives. Hence, if g is low, the equilibrium must involve some publishing and some deterrence of wrongdoing.

Similarly, the equilibrium where type L never does wrong will not exist if g is too large, as this would put J_1 off publishing. To analyze this more formally, J_1 prefers publishing if and only if

$$p + \frac{(g\omega_1^H + (1-g)\omega_1^L\lambda_1^L)\ell_1^J + (g(1-\omega_1^H) + (1-g)(1-\omega_1^L)\lambda_0^L)\sigma\ell_0^J}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)(\omega_1^L + (1-\omega_1^L)\sigma)} \geq 0. \tag{11}$$

As g approaches 1, the left-hand side of (11) becomes $p + \omega_1^H\ell_1^J + (1-\omega_1^H)\ell_0^J$, which is negative due to our assumption of $p < -\ell_1^J$. Hence, J_1 will not publish whenever g is close to 1, so that both types of public figure will go undeterred from wrongdoing.

The following proposition formally establishes that L always (never) does wrong if g is close to one (zero), and that the existence of these equilibria is monotonic in g in the sense that they exist below or above unique thresholds for g:

Proposition 2 *(i) In equilibrium, L does no wrong in the first period and J_1 publishes if and only if g is below a threshold $\underline{g}$, and L does wrong in the first period and J_1 does not publish if and only if g is above a threshold $\overline{g}$. For $g \in [\underline{g}, \overline{g}]$, L and J_1 randomize.*

(ii) $\underline{g}$ and $\overline{g}$ are nondecreasing in p, ℓ_0^J and ℓ_1^J. Furthermore, $\underline{g}$ is weakly decreasing in ℓ_0^L and constant in ℓ_1^L, and $\overline{g}$ is weakly decreasing in ℓ_1^L and constant in ℓ_0^L.[21]

Proof. See the Appendix. ■

Again, J_1's equilibrium publication decision is inversely related to the veracity of the story: J_1 publishes in a parameter range where L does no wrong, so that the story is false whenever the public figure is type L. Furthermore, we don't observe any lawsuits if $g > \overline{g}$, as J_1 never

[21] The comparative statics for $\underline{g}$ are only local for given wrongdoing choices of type H, i.e. $\omega_1^H = 1$ or $\omega_1^H = 0$.

publishes in this case, whereas type L never does wrong but sues with some probability if $g < \underline{g}$ and $-\ell_0^L < b + \sigma r$. Hence, the cases that actually go to court are not a random selection of all situations where a journalist obtains evidence but rather skewed towards the subset of these situations in which the story is false.

Part (ii) of Proposition 2 provides some comparative statics of the thresholds for g below (above) which L never (always) does wrong in equilibrium. In order to understand the mechanisms behind these results, let us construct $\underline{g}$ and $\overline{g}$ for the special case where (7) holds, in which case type H always does wrong irrespective of the media's publication strategy ($\omega_1^H = 1$).[22] Consider first an equilibrium where L never does wrong and J_1 always publishes ($\omega_1^L = 0$ and $\eta_1 = 1$). If $-\ell_0^L > b + \sigma r$, Proposition 1 has shown that L never sues in the first period ($\lambda_0^L = \lambda_1^L = 0$). Substituting for these equilibrium strategies in (11) and solving for g implies that J_1 will publish if and only if $g < \frac{\sigma p}{-\ell_1^J - (1-\sigma)p}$.

If $-\ell_0^L \leq b + \sigma r$, L prefers suing in the first period when doing so would make J_2 believe that she was type H for sure. Of course, this will be taken into account by J_2 when updating his beliefs μ_2^S after observing a first-period lawsuit. The equilibrium is in mixed strategies, with L suing with probability λ_0^L. This requires that J_2 publishes with exactly the right probability after observing a first-period lawsuit to ensure that L is exactly indifferent between suing and not suing, which would again identify herself as type L. J_2, in turn, will only randomize with the required probability if $\mu_2^S \in \{\widetilde{\mu}^1, \widetilde{\mu}^2\}$, in which case there is a continuum of equilibrium publication probabilities. If type L's litigation costs $-\ell_0^L$ are very low, L needs a low second-period profit $\pi_2^L(\mu_2^S)$ to make her indifferent between suing and not suing in the first period, so that $\mu_2^S = \widetilde{\mu}^1 = \frac{\sigma p}{-\ell_1^J - (1-\sigma)p}$. For larger $-\ell_0^L$, but still below $b + \sigma r$, we have $\mu_2^S = \widetilde{\mu}^2 = \frac{p}{-\ell_1^J}$.[23] Recalling that consistent beliefs μ_2^S are given by (9), this uniquely defines L's equilibrium litigation probability λ_0^L, which then can be used to substitute in (11). Doing so yields the threshold for g below which L never does wrong and J_1 always publishes in the case where (7) holds:

$$\underline{g} = \begin{cases} \frac{\sigma p}{-\ell_1^J - (1-\sigma)p - \ell_0^J \frac{-\ell_1^J - p}{\sigma p}}, & \text{if } -\ell_0^L \leq \frac{\sigma}{1-\sigma}[(1-\sigma)r - b]; \\ \frac{\sigma p}{-\ell_1^J - (1-\sigma)p - \ell_0^J \frac{-\ell_1^J - p}{p}}, & \text{if } \frac{\sigma}{1-\sigma}[(1-\sigma)r - b] < -\ell_0^L \leq b + \sigma r; \\ \frac{\sigma p}{-\ell_1^J - (1-\sigma)p}, & \text{if } -\ell_0^L > b + \sigma r. \end{cases} \tag{12}$$

This threshold is smoothly increasing in p and ℓ_1^J and, if $-\ell_0^L \leq b + \sigma r$, in ℓ_0^J. Of course, a higher publication benefit makes the equilibrium in which J_1 always publishes more likely. On the downside, J_1 might get sued by type H, who had done wrong, thus imposing a negative litigation payoff ℓ_1^J on him. The larger this litigation payoff (i.e., the smaller this litigation costs) is for J_1, the less important is type H's prior probability in preventing the equilibrium in which J_1 always publishes. If $-\ell_0^L \leq b + \sigma r$, L will also sometimes sue, so that the same

[22] In an earlier working paper version Acheson and Wohlschlegel (2018b), we provide a full characterization of equilibrium for this case.

[23] Recall that these thresholds $\widetilde{\mu}^1$ and $\widetilde{\mu}^2$ where derived in the discussion before Lemma 1.

reasoning is true for ℓ_0^J in this case.

Furthermore, as $-\ell_0^L$ increases, equation (12) may transition into another case, which means that $\underline{g}$ jumps discontinuously upwards. This is because L will sue with a lower probability in the first period as doing so becomes more costly for her. This lower litigation probability, in turn, makes publishing less costly for J_1, thus making the equilibrium where he always publishes more likely.

The threshold value $\overline{g}$ above which J_1 never publishes can be constructed in a similar way: In an equilibrium where L always does wrong and J_1 never publishes, L's litigation probability λ_1^L is determined by $\mu_2^S \in \{\widetilde{\mu}^1, \widetilde{\mu}^2, 1\}$, depending on whether $-\ell_0^L > b + \sigma r$ or moderately or much below that threshold, respectively. Again, using this to substitute for λ_1^L in (11) yields $\overline{g} = \frac{p}{-\ell_1^J}\mu_2^S$, which is equal to

$$\overline{g} = \begin{cases} \frac{\sigma p}{-\ell_1^J - (1-\sigma)p}\frac{p}{-\ell_1^J}, & \text{if } -\ell_1^L \leq \frac{\sigma}{1-\sigma}[(1-\sigma)r - b]; \\ \left(\frac{p}{-\ell_1^J}\right)^2, & \text{if } \frac{\sigma}{1-\sigma}[(1-\sigma)r - b] < -\ell_1^L \leq b + \sigma r; \\ \frac{p}{-\ell_1^J}, & \text{if } -\ell_1^L > b + \sigma r. \end{cases} \tag{13}$$

As in the previous case, higher p and ℓ_1^J make publishing more attractive and the equilibrium where J_1 never publishes less likely (in which both types of public figure do wrong). Furthermore, higher litigation costs for type L make L's litigation probability jump to lower levels and, thus, make publishing less costly for J_1 and $\overline{g}$ jump upwards.

This impact of type L's litigation cost $-\ell_k^L$ on equilibrium wrongdoing and publishing is a unique feature of the two-period model: In the single-period model, which we have argued before is equivalent to the discussion in Section 3, type L always sues if $\ell_k^L > 0$, which completely discourages the journalist from publishing, and L never sues if $\ell_k^L \leq 0$, which implies the equilibrium characterized in Lemma 1. Intuitively, journalists in later periods can tell from previous litigation choices whether public figures are more or less likely to sue, which means low future payoffs for public figures who let publications go unchallenged early on. However, imitating libel bullies by suing journalists in early periods can alleviate this problem, which may also make wrongdoing more attractive. Hence, type L's wrongdoing incentives in the first period depend on her litigation cost.

Another consequence of this equilibrium feature is, of course, the direct effect of more litigation. For large g, this does not make any difference, as J_1 will not publish in this case anyway. For small g, if L is deterred from wrongdoing and J_1 always publishes, lower litigation costs increase the probability that L sues. As the story is, in equilibrium, wrong if the public figure is of type L, such a lawsuit will improve the information disseminated to the general public.

Last, type L's litigation costs $-\ell_k^L$ also impact on second-period equilibrium play: If $-\ell_k^L$ increases but stays below $b + \sigma r$, J_2 must publish with a higher probability in order to make L indifferent between suing and not suing in the first period. This makes wrongdoing in the later period more likely to be detected, but also an honest type-L public figure more likely to

be defamed.

Another interesting comparison between the two-period and the single-period models concerns the chilling effect. Recall from Section 3 that, in the single-period model, the journalist is completely deterred from publishing if and only if $g > \frac{p}{-\ell_1^J}$. This result carries over to the two-period model in the case where $-\ell_0^L > b + \sigma r$, i.e., where L never wants to sue in the first period so that both models are equivalent. However, if $-\ell_0^L \leq b + \sigma r$, the range of g where the first-period journalist is deterred from publishing widens considerably due to L's incentives to imitate the libel bully H by suing for defamation.

Note that the parameter range defined in Assumption 1, on which the analysis so far has been based, may not always be the most plausible one. Hence, we will now briefly discuss equilibrium outside this range. If type H's litigation payoff falls below zero, type L's incentives to imitate her go down. However, as long as there are some incentives for type H to sue for defamation (i.e., if $\ell_1^H < 0 < \ell_0^H$), this is all we need for our result of type L imitating type H to go through. If even $\ell_0^H < 0$, neither type of public figure would ever sue, which the journalist will anticipate and always publish. On the other hand, if L's litigation payoff becomes positive, her incentives to sue increase and the journalist's incentives to publish go down. If even $\ell_1^L > 0$, both types of public figure will always sue and the journalist will never publish.

Similarly, one might think that the assumption that the journalist's litigation costs outweigh the publication gain is extreme. If even ℓ_0^J increases above $-p$, the journalist prefers publishing even when anticipating that he will definitely be sued. In this case, there is no way to deter the journalist from publishing, regardless of whether the story is false. In other words, this shows that if one thinks that the argument of imitating libel bullies is relevant in some cases and may cause a chilling effect by discouraging even the publication of true stories, then one must also think that the assumption $\ell_1^J < -p$ holds in those cases at least.

5 Legal Reform

In this section, we analyze the effects that potential legal reforms might have on the incentives to do wrong, publish and litigate in equilibrium. As in Garoupa (1999a), four scenarios are possible: (i) full honesty, where the public figure does no wrong and no wrongdoing is alleged by the media; (ii) exposed dishonesty, where the public figure does wrong and is exposed by the media; (iii) hidden dishonesty, where the public figure does wrong but is not exposed by the media; and (iv) defamation, where wrongdoing does not take place but is alleged by the media. With regards to welfare comparisons of these scenarios, we follow Garoupa (1999a) by assuming that full honesty is the socially optimal scenario and that exposed dishonesty is socially preferred to hidden dishonesty but not imposing any other preferences between scenarios.[24]

This paper is concerned with a situation in which there are libel bullies who cannot be

[24]Note, however, that there seems to be a broad consensus in the US legal literature that, in general, errors in (iii) are more socially harmful than errors in (iv) (e.g. Schauer (1978), p. 688).

deterred from suing for defamation even if the allegation made by the media is true, and about incentives to imitate these libel bullies in order to discourage media reports. We therefore focus on the impact of defamation law on the strategy of the type of public figure that could potentially be deterred from suing for libel and from doing wrong (type L) in the period in which these imitation incentives might exist (i.e., the first period) and, of course, on the media's strategy. That is to say, we will refer to the equilibrium where L does (does not) do wrong in the first period and J_1 does not (does) publish as hidden dishonesty (defamation).[25]

The first thing to note is that there are no parameter constellations such that the equilibrium always exhibits full honesty: According to Proposition 2, in any equilibrium where L never does wrong, the journalist always wants to publish whenever possible. A situation in which type L does not do wrong and the journalist does not publish despite observing evidence is only possible with some probability within the mixed strategy equilibria that exist for intermediate values of g. Outside the parameter range of Assumption 1 that this proposition (and the analysis in Sections 3 and 4) is confined to, there is no pure-strategy full-honesty equilibrium either: Leaving that parameter range either establishes defamation (if $\ell_0^J > -p$ or $\ell_0^H < 0$) or hidden dishonesty (if $\ell_1^L > 0$ or $b \geq (1-\sigma)r$) as equilibrium for any g.

As there is no clear social optimum, we shall discuss instead the impact of several potential ways to reform defamation law on equilibrium. Note that equilibrium also depends on parameters beyond the legislator's control, which is why we will assess potential legal reforms with regards to the impact they have on the parameter ranges that support certain equilibria. More specifically, we will focus on the sets of prior probabilities g for the public figure to be a true libel bully (type H) that support the pure-strategy equilibria of defamation and of hidden dishonesty, respectively. In our model, the legal environment is represented by variables ℓ_k^i, which cannot be directly manipulated by a legislator. Rather, policy variables like the size of damages, the accuracy of defamation courts or the size and allocation of legal costs can impact on these ℓ_k^i in different ways. In order to analyze the impact of these variables on equilibrium, we need to be more specific about ℓ_k^i.

Let a type-i public figure's net payoff from litigation, depending on whether she has done wrong ($k = 1$) or not ($k = 0$), be parametrized as

$$\ell_k^i := q_k(\alpha + \beta_i) - (1 - q_k)\widetilde{c} - c_P = \left[\frac{1}{2} + \rho\left(\frac{1}{2} - k\right)\right](\alpha + \beta_i + \widetilde{c}) - (c + \widetilde{c}), \tag{14}$$

where α is the damages that a journalist will have to pay to the public figure when losing at trial, β_i a type-i public figure's indirect benefit from being vindicated at trial,[26] c_P the part of legal cost of a defamation suit that the plaintiff needs to bear irrespective of the outcome, $\widetilde{c}$ that part of the cost that only the losing litigant bears, and q_k the probability of the public figure

[25] Therefore, a publication in what we label a 'defamation' equilibrium might well be true and, thus, socially beneficial, if the public figure is type H.

[26] More generally, α can be seen as a zero-sum part of litigation payoffs that benefits the public figure to exactly the same extent as it harms the journalist, and β_i is the plaintiff-specific benefit of winning at trial. Such a distinction between these types of litigation payoffs has also been made by Garoupa (1999a).

winning the defamation lawsuit.[27] We assume that this probability depends on whether the journalist's allegation is true ($k = 1$), in which case the public figure wins with probability $\frac{1-\rho}{2}$, or false ($k = 0$), in which case the public figure wins with probability $\frac{1+\rho}{2}$. This parametrization permits us to conveniently interpret $\rho \in [0, 1]$ as the defamation court's accuracy: If the court is perfectly accurate ($\rho = 1$), the plaintiff wins if and only if the allegation is false, whereas a completely inaccurate court ($\rho = 0$) would just be a coinflip independent of the truth of the journalist's allegations.

Similarly, let us parametrize the journalist's litigation payoff as

$$\ell_k^J := -q_k(\alpha + \beta_J + \widetilde{c}) - c_J = -\left[\frac{1}{2} + \rho\left(\frac{1}{2} - k\right)\right](\alpha + \beta_J + \widetilde{c}) - c_J. \tag{15}$$

The journalist pays damages α and the success-dependent part $\widetilde{c}$ of legal cost, and suffers defendant-specific non-monetary costs β_J, after losing at trial, which occurs with probability q_k, and pays non-contingent legal cost c_J.

With these definitions, we will now discuss some potential avenues for legal reform in more detail. Recall that the comparative statics with respect to $\underline{g}$ presented in Proposition 2 (ii) were local for a given choice of ω_1^H, so that we will here confine the discussion to reforms that are so small that they will not change type H's wrongdoing choice in the first period.

Let us start by analyzing the impact of making the damages awarded against a losing defendant more generous. A \$1 increase in damages awarded to a victorious plaintiff increases a type-i public figure's expected payoff from litigation ℓ_k^i by the probability of winning, q_k, and reduces a journalist's ℓ_k^J by the same amount, which depends on whether the journalist's allegation was true ($k = 1$). All of these changes reduce $\underline{g}$ and $\overline{g}$, i.e., make the defamation equilibrium less likely and the hidden-dishonesty equilibrium more likely.

A similar effect is implied when changing the substantive tests used to determine the outcome of defamation cases so that the chances of success in court are made more favorable to one or other of the parties. For example, the rules set out in NYT v Sullivan make it much less likely that public figure plaintiffs will succeed with a defamation suit in the US. In terms of the above definitions, this means in a reduction of q_k and, therefore, an increase in ℓ_k^J and a reduction in ℓ_k^i, for both $k = 0$ and $k = 1$. In this sense, our model suggests that, by making the law more favorable to defendants, the Sullivan case has made the defamation equilibrium more and the hidden-dishonesty equilibrium less likely; and that reforms which 'open up' libel laws to make it easier for plaintiffs to succeed would have the opposite effect.

Another way for a legislator to influence litigation, publication and wrongdoing incentives is to invest in the accuracy ρ of defamation courts. Of course, this is only possible up to a certain limit, which is a major cause of the uncertainty of litigation outcomes that helps drive the chilling effect in the first place (e.g. Schauer (1978), pp. 687-88). The accuracy in the sense

[27] To get an idea of that probability in practice, in a study of US defamation litigation in the 1980s and 1990s, Logan (2001) reported that around 85% of claims filed by public figure plaintiffs were dismissed before trial. Plaintiffs were awarded damages in around 65% of cases that reached trial; but many of those verdicts were subsequently reversed or reduced on appeal.

of how relevant the truth or falsity of the statement is to the outcome of lawsuits relates to one of the criticisms of the focus on the defendant's fault in US defamation law since NYT v Sullivan (in e.g. Bezanson (1988), pp. 541-42 and 554-55). In this sense, the Sullivan rules, by focusing on the defendant's conduct, also decrease the accuracy of courts' decisions.[28]

Equations (14) and (15) imply that a higher accuracy increases the litigation payoff ℓ_0^i and ℓ_1^J if the litigant is in the right, and reduces it otherwise, i.e., it reduces ℓ_1^i and ℓ_0^J. With Proposition (2), therefore, improving accuracy of defamation courts has mixed effects: The effect on the type L public figure's litigation payoff reduces $\underline{g}$ and increases $\overline{g}$, i.e., it makes both the defamation and the hidden-dishonesty pure strategy equilibria less likely and a mixed strategy equilibrium more likely. However, the effects on the journalist's litigation payoff work in opposite directions, as both $\underline{g}$ and $\overline{g}$ are nondecreasing, and within some parameter ranges strictly increasing in ℓ_1^J and ℓ_0^J. Therefore, general conclusions cannot be drawn. However, in the example discussed above, where (7) holds, $\overline{g}$ does not depend on ℓ_0^J, so that the effect of an improved accuracy in increasing ℓ_1^J also increases $\overline{g}$. Hence, the effect of accuracy on $\overline{g}$ in this special case is unambiguously positive. Furthermore, noting that the effect of ρ on ℓ_1^J and ℓ_0^J have the same absolute size, it can be shown that the effect of improved accuracy on the journalist's litigation payoff also increases $\underline{g}$, which is countervailing to the aforementioned effect via type L's litigation payoff, and it cannot be generally established which effect dominates. When considering small changes in the court's accuracy, one needs to bear in mind that the thresholds $\underline{g}$ and $\overline{g}$ are constant in the public figure's litigation payoff within certain intervals, as their relationship is step-wise. Hence, if the cutoffs for these steps are not reached, the only noticeable effect will be that via the media's litigation payoff.

Next, let us consider the impact of changing the size and the allocation of legal costs of litigation. Increasing the legal costs that a litigant incurs irrespective of the outcome of the lawsuit reduces both the public figure's litigation payoff, which increases $\underline{g}$ and $\overline{g}$, and the journalist's litigation payoff, which reduces $\underline{g}$ and $\overline{g}$. Again, recall that the former effect is step-wise, whereas the latter is smooth. Similarly, increasing the part of the costs that the losing litigant bears, $\widetilde{c}$, reduces both litigants' expected litigation payoff, with similar effects on equilibrium. However, an interesting avenue for legal reform might be to keep litigants' overall court costs constant but make them more dependent on winning by increasing $\widetilde{c}$ but reducing c_P and c_J at the same time or, in other words, moving the cost allocation from the US rule towards the English rule. In that case, ℓ_0^i and ℓ_1^J increase, whereas ℓ_1^i and ℓ_0^J go down, so that the effect is similar to that of an improvement in accuracy.[29]

In summary, there are two types of legal reform: One type moves equilibrium generally in one of the litigants' favour by making a pure-strategy defamation equilibrium more and a pure-

[28] At the same time, one could argue that the US rules skew the errors courts make in favour of defendants. While the effect of this asymmetry is similar to the above discussion of increasing damages awards, the following discussion focuses on symmetric changes in accuracy.

[29] This beneficial incentive effect is often referred to by advocates of the English rule such as Hollander (1989) and Windon (2010).

strategy hidden dishonesty equilibrium less likely or *vice versa*. This type can straightforwardly be implemented, for instance, by increasing or decreasing the damages that a losing defendant must pay to the plaintiff. The other type of reform moves the thresholds $\underline{g}$ and $\overline{g}$ in opposite directions, thus making both pure-strategy equilibria either more or less likely. Under certain parameter constellations, this may be the case when changing the courts' accuracy or the fraction of costs to be borne by the losing litigant.

The main contribution of this paper is to take into account the possibility that a public figure wants to imitate libel bullies in order to discourage future publications. Therefore, it is useful to analyze the effect that this possibility has on the impact of legal reform. As discussed above, the most striking difference to a one-period model in which this kind of signalling is not feasible is that, as long as type L does not directly benefit from going to court, her litigation payoff does not affect the range of priors in which the pure-strategy defamation or hidden dishonesty equilibria emerge. Hence, the qualitative impact of a change in the size of damages is the same in both models, as type L's litigation payoff works in the same direction as the journalist's in the two-period model anyway.

However, an important difference emerges whenever the effect via type L's litigation payoff works in the opposite direction than the journalist's. For instance, when making court fees more expensive for both litigants at the same time, the effects of both of these changes in litigation costs impact on the thresholds $\underline{g}$ and $\overline{g}$ in opposite directions. On the other hand, by taking this difference into account legal reform might, under certain parameter constellations, be used to make both undesirable pure-strategy equilibria, defamation and hidden dishonesty, less likely, such as by increasing the accuracy or the fraction of legal costs borne by the losing litigant.

6 Conclusion

We have analyzed a two-stage model of defamation law with endogenous wrongdoing, publishing and litigation decisions, taking into account the fact that it becomes public knowledge among journalists over time which public figures are more or less likely to sue for defamation when exposed for alleged wrongdoing by the media. We assume that there are two types of public figures, one of which benefits, in expectations, from suing for defamation, whereas the other type incurs a net cost from a defamation lawsuit. In equilibrium, the latter type will imitate the litigious type in the first period with some probability as long as her litigation loss is not too high. This also makes the first-period journalist less likely to publish and, therefore, the non-litigious type of public figure more likely to do wrong in the first period in equilibrium. Thus, our model exhibits libel bullying – defamation lawsuits brought solely to discourage future media stories. The model also shows the problem that publication decisions are mainly based on the media's beliefs of how likely a public figure is to sue, although less litigious public figures also have less incentive to do wrong in the first place, which makes a journalist's evidence of their wrongdoing less likely to be accurate.

Due to these incentives to litigate in order to manipulate future journalists' beliefs, equilibrium wrongdoing and publication choices in the earlier period depend on the otherwise non-litigious type's litigation payoffs. As litigation gets more costly for the public figure, the equilibrium where she does (does not) do wrong and the media do not publish (do publish) becomes less (more) likely. This may create a countervailing effect to legal reform that a legislator can use to make both of these socially undesirable pure strategy equilibria less likely, albeit at the cost of reducing access to justice for defamation victims.

Beyond the topic of defamation law, our model can also be used to analyze the impact of whistleblower protection law, which is aimed at mitigating the deterrent effect that punishment of whistleblowers by employers and co-workers has on insiders' willingness to expose wrongdoing. When interpreting our model in such a way, the parameters ℓ_w^i correspond to the employer's and the whistleblower's expected payoff when the former punishes the latter, which might involve the case going to some employment tribunal. Both players' expected payoffs from such an action may be negative, but an exposed employer may still be willing to incur that cost in order to deter future whistleblowing. Furthermore, Mechtenberg, Muehlheusser, and Roider (2020), for instance, argue that being too tough on employers who dismiss a whistleblower may create incentives for low productivity employees to make allegations that they know are false just in order to gain whistleblower status in an attempt to secure their job. In this setting, our model implies that the employer's incentives to punish a whistleblower in discord with whistleblower protection law is decreasing in her expectations of future whistleblowers' incentives to blow the whistle. The policy recommendation from our model is, therefore, to focus whistleblower protection efforts on increasing the gains from whistleblowing rather than on imposing higher costs on employers who dismiss or demote whistleblowers.

In our model, the second-period journalist J_2 uses two first-period signals to update his beliefs on the public figure's type: the question of whether or not the first-period journalist J_1 has published an allegation on this public figure, and the question of whether the public figure has taken J_1 to court over this publication. In reality, more detailed information on whether the public figure has done wrong in the first period and, therefore, on her type, might be revealed in the first period. For instance, if there is a lawsuit in the first period, the outcome of this lawsuit might become public knowledge. As public figures who have actually done wrong are less likely to prevail in court, the trial outcome is an informative signal for the public figure's first-period wrongdoing decision and, if type L is less likely to do wrong in equilibrium than type H, for her type. Including such a more detailed signal in the model will cause J_2's strategy to depend on more information sets but would not change the qualitative results. As such an extended model would introduce more case distinctions (regarding the way in which the trial outcome depends on the public figure's wrongdoing) and, therefore, would come at a considerable loss of expositional clarity, we decided to go with the simplifying assumption that the trial outcome is private knowledge to the litigants.

Similarly, the first-period case may be settled rather than going to court. When taking

this possibility into account, the important difference to our version of the model is the way in which the second-period journalist J_2 updates his beliefs on the public figure's type if a first-period case is settled privately. For instance, in an equilibrium in mixed strategies, in which J_2 publishes with such a probability in the second period to make type L indifferent between filing suit and not doing so, type H will still prefer filing suit to settling out of court. Therefore, J_2 will believe the public figure to be type L with certainty whenever there was no court case in the first period. Hence, the Bayesian updating after out-of-court settlement would work in exactly the same way as it does in our model if suit is not filed in the first period. Furthermore, given a settlement amount and J_2's beliefs that make type L indifferent between settling and filing suit, slightly increasing the settlement amount makes type L strictly prefer settling. Hence, the only settlement amount that can be supported in an equilibrium will be the one that makes J_1 indifferent between settling and being taken to court. This implies that J_1's decision of whether to publish in the first place, and, therefore, the public figure's decision of whether to do wrong in the first period, will depend on the same parameters in a similar way as in our model. Again, we decided to rule out this complication of the model.

Another possible extension of the model is to allow for the possibility that the same journalist is active in both periods. This would require two kinds of changes to the model: First, we would have to assume that, in the second period, the journalist knows the first-period trial outcome. Second, publishing a story in the first period gives rise to an additional benefit for the journalist: If the public figure takes the journalist to court in the first period, the trial outcome can be used by the journalist to update his beliefs on the public figure's type. However, the basic insight of the model, that a type L public figure may want to bring negative expected value lawsuits in order to appear to be type H would remain unaffected.

Our analysis is driven by the assumptions that there are some public figures who directly benefit from bringing a defamation lawsuit and others who incur net costs, and that journalists never want to publish if they know for sure that they are going to be sued if they do. We make these assumptions because, as we have argued in the Introduction, these cases are generally plausible, and because it is under these assumptions that the possibility to imitate a libel bully makes a real difference in equilibrium. For instance, if both types of public figure's litigation payoffs have the same sign, there is no point in imitating one another. Similarly, if journalists' costs of being sued for defamation are so small that they always prefer publishing, there is no point in manipulating their beliefs as this would not change their actions. In all of these cases, the equilibrium would be either one of hidden dishonesty or defamation for all prior probabilities of types of public figure.

Recently, Arbel and Mungan (2019) have criticized the assumption of exogenous reputation losses for public figures from wrongdoing allegations that all of the existing economic literature on defamation law makes. They argue that a rational audience will take the equilibrium probability into account that the media publish a false story. For instance, if more defamatory publications are deterred in equilibrium under a stricter defamation law, the general public

will take stories that are actually published more seriously, which means that those defamatory claims that go undeterred will incur a larger reputation loss for the public figure.

We disregard this effect as the role of the media in a model with a completely rational audience would be very different than in our model: In our model, the media's role is to filter evidence, and their prior about the veracity of that evidence is no more accurate than that of the general public. The only difference is that the media engage in Bayesian updating based on equilibrium strategies, whereas we assume that the general public cannot. By contrast, the media in a model of a Bayesian general public would have to have an informational advantage regarding the veracity of the evidence in order to be meaningful at all. As we think that both roles, of filtering evidence and confirming its veracity, are important in reality, we consider both approaches to be complementary and leave the analysis of libel bullies in a model of endogenous reputation effects for future research.

Appendix

A Proof of Lemma 1

For the case where (7) holds, we have already proven the Lemma with $\widetilde{\mu}^1 = \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}$ and $\widetilde{\mu}^2 = \frac{p}{-\ell_1^J}$. If $\ell_1^H < \sigma\ell_0^H+(1-\sigma)r-b$, type H will refrain from doing wrong if $\eta_2 > \frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}$. Using (4), both types not doing wrong and J_2 publishing is an equilibrium if and only if $\mu_2 \leq \frac{p}{-\ell_0^J}$, and both types doing wrong and J_2 not publishing is an equilibrium if $\mu_2 \geq \frac{p}{-\ell_1^J}$. For beliefs in between these boundaries, the equilibrium will be in mixed strategies.

If $0 < \sigma\ell_0^H < \ell_1^H < \sigma\ell_0^H + (1-\sigma)r - b$, type H will want to do wrong for some mixed strategies of J_2 for which L would still prefer not doing wrong. Hence, there are some beliefs μ_2 for which type L is indifferent between doing wrong and not doing wrong, whereas H strictly prefers doing wrong. With (4), this can be an equilibrium if $\mu_2 \in \left[\frac{\sigma p}{-\ell_1^J-(1-\sigma)p}, \frac{p}{-\ell_1^J}\right]$. Hence, if $\mu_2 = \frac{p}{-\ell_0^J}$, any $\eta_2 \in \left[\frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}, 1\right]$ may be an equilibrium with L not doing wrong and H randomizing, and for $\mu_2 = \max\left\{\frac{p}{-\ell_0^J}, \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}\right\}$, any $\eta_2 \in \left[\frac{b}{(1-\sigma)r}, \frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}\right]$ may be an equilibrium with H doing wrong and L randomizing. If $\left(\frac{p}{-\ell_0^J}, \frac{\sigma p}{-\ell_1^J-(1-\sigma)p}\right)$ is non-empty, then for any μ_2 in this open interval, $\eta_2 = \frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}$, L does no wrong and H randomizes.

If $0 < \ell_1^H \leq \sigma\ell_0^H$, type H's wrongdoing incentives are weaker than type L's, i.e., type L will want to do wrong for some mixed strategies of J_2 for which H would still prefer not doing wrong. Hence, for $\mu_2 = \frac{p}{-\ell_0^J}$, any $\eta_2 \in \left[\frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}, 1\right]$ may be an equilibrium, whereby H does no wrong and L randomizes if $\eta \in \left[\frac{b}{(1-\sigma)r}, 1\right]$, and L does wrong and H randomizes if $\eta_2 \in \left[\frac{b}{(1-\sigma)r-\ell_1^H+\sigma\ell_0^H}, \frac{b}{(1-\sigma)r}\right]$.

B Proof of Proposition 2

Consider an equilibrium in which J_1 always publishes, and L does not *strictly* prefer doing wrong. As $\ell_1^H \geq 0$, H always sues. This implies $\lambda_1^L < 1$ as, if L always sued in the first period, J_1 would always be sued by either type, so that J_1 would prefer not to publish.

We start by showing that type L prefers not doing wrong in this situation. In general, type i prefers doing wrong if and only if

$$b - r(1-\sigma) \geq -\max\{\ell_1^i + \pi_2^i(\mu_2^S), \pi_2^i(\mu_2^N)\} + \sigma \max\{\ell_0^i + \pi_2^i(\mu_2^S), \pi_2^i(\mu_2^N)\} + (1-\sigma)\pi_2^i(\mu_2^{NP}) \quad (16)$$

For any μ_2, Lemma 1 implies that $\pi_2^L(\mu_2) \geq -\sigma r$, so that this is also true for μ_2^{NP} (last term on the right-hand side of (16)), and also for μ_2^N, so that $\max\{\ell_0^L + \pi_2^L(\mu_2^S), \pi_2^L(\mu_2^N)\} \geq -\sigma r$. Furthermore, as $\lambda_1^L < 1$, $\max\{\ell_1^L + \pi_2^L(\mu_2^S), \pi_2^L(\mu_2^N)\} = \pi_2^L(\mu_2^N) = -\sigma r$, where the second equation is due to the fact that H always sues if $0 \leq \ell_1^H$, so that $\mu_2^N = 0$. Hence, the right-hand side of (16) is at least as large as $\sigma r - \sigma^2 r - (1-\sigma)\sigma r = 0$, whereas the left-hand side is negative by assumption. Hence, L never does wrong.

The next step is to show that whether H wants to do wrong does not depend on g. Due to (16), when J_1 is known to publish and L not to do wrong, H prefers doing wrong in the first period if and only if

$$\pi_2^H(\mu_2^{NP}) \leq \pi_2^H(\mu_2^S) - r + \frac{b + \ell_1^H - \sigma\ell_0^H}{1-\sigma} \quad (17)$$

If, in equilibrium, $\omega_1^H = 1$, then $\mu_2^{NP} = 0$, whereas in an equilibrium with $\omega_1^H = 0$, $\mu_2^{NP} = g$. Furthermore, recall from Lemma 1 that $\pi_2^i(\cdot)$ are weakly increasing functions. Hence, the left-hand side of (17) is weakly larger in an equilibrium with $\omega_1^H = 0$ than in one with $\omega_1^H = 1$. As for the right-hand side of (17), $\mu_2^S = \frac{g(\omega_1^H + (1-\omega_1^H)\sigma)}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)\sigma\lambda_0^L}$, which is larger for $\omega_1^H = 1$ than for $\omega_1^H = 0$ for any given $\lambda_0^L > 0$ and equal for $\lambda_0^L = 0$. Hence, whenever (17) is violated for $\omega_1^H = 1$, implying that $\omega_1^H = 1$ cannot be an equilibrium, it must also be violated for $\omega_1^H = 0$, implying that $\omega_1^H = 0$ can be an equilibrium. Last, whether (17) is satisfied (and, thus, whether $\omega_1^H = 1$ in equilibrium) depends on b, ℓ_0^H, ℓ_1^H, σ, $\pi_2^H(0)$, which, in turn, only depends on the aforementioned parameters, and $\pi_2^H(\mu_2^S)$.

This last expression, $\pi_2^H(\mu_2^S)$, is equal to b if $\lambda_0^L = 0$ (as $\mu_2^S = 1$ in this case), the condition for which is, according to Proposition 1, $-\ell_0^L \geq b + \sigma r$. If, on the other hand, $-\ell_0^L < b + \sigma r$, then $\lambda_0^L > 0$, so that, in equilibrium, $\pi_2^L(\mu_2^S)$ must be such that L is indifferent between suing and not suing in the first period. According to Lemma 1, L's second-period payoff $\pi_2^L(\mu_2^S)$ depends on J_2's publishing probability η_2 and, potentially, on b, r and σ. Hence, in order to make L indifferent between doing and not doing wrong, η_2 must be chosen to equate both sides of (16), so that the equilibrium η_2 will depend on the other elements of this equation, none of which depends on g. As a conclusion, whether ω_1^H is, in equilibrium, equal to 1 or 0 does not depend on g, so that we can treat ω_1^H as exogenous in the following.[30]

[30]Note that we have not technically ruled out that H might choose mixed strategies in some other equilibrium that might exist along with those just specified, but for the sake of simplicity we forego this discussion by assuming

Given some exogenous $\omega_1^H \in \{0,1\}$, we will now show that whether J_1 indeed prefers publishing depends monotonically on g, i.e., J_1 prefers publishing (in which case the situation explained above is indeed an equilibrium) if and only if g is below some threshold $\underline{g}$. Consider first the case in which $-\ell_0^L \geq b + \sigma r$, which, according to Proposition 1, implies that $\lambda_0^L = 0$. When substituting for $\lambda_0^L = 0$ and $\omega_1^L = 0$, the left hand side of (11) is equal to

$$p + \frac{g(\omega_1^H \ell_1^J + (1-\omega_1^H)\sigma\ell_0^J)}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)\sigma}, \tag{18}$$

the derivative of which w.r.t. g is $\frac{\sigma(\omega_1^H \ell_1^J + (1-\omega_1^H)\sigma\ell_0^J)}{[g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)\sigma]^2} < 0$. Furthermore, the expression in (18) is $p > 0$ if $g = 0$ and $p + \frac{\omega_1^H \ell_1^J + (1-\omega_1^H)\sigma\ell_0^J}{\omega_1^H + (1-\omega_1^H)\sigma} < p + \ell_1^J < 0$ if $g = 1$. Hence, there exists an $\underline{g}$ such that, in equilibrium, L does no wrong and J_1 publishes if and only if $g \leq \underline{g}$.

Let us now turn to the case where $-\ell_0^L < b + \sigma r$, for which Proposition 1 implies $0 < \lambda_0^L < 1$. As explained above, this requires J_2 to publish with exactly that probability η_2 that makes L indifferent between suing and not suing in the first period. For J_2 to randomize, in turn, he needs to be indifferent between publishing and not publishing, which requires a specific second-period belief $\widetilde{\mu}_2$ (see Lemma 1). Depending on the first-period strategies (and again taking into account that $\omega_1^L = 0$), the second-period beliefs after observing a lawsuit in the first period are

$$\mu_2^S = \frac{g(\omega_1^H + (1-\omega_1^H)\sigma)}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)\lambda_0^L\sigma} \tag{19}$$

Equating $\mu_2^S = \widetilde{\mu}_2$ and solving for λ_0^L yields

$$\lambda_0^L = \frac{g}{1-g}\frac{\omega_1^H + (1-\omega_1^H)\sigma}{\sigma}\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}. \tag{20}$$

Using (20) to substitute for λ_0^L on the left-hand side of (11) yields

$$p + g\frac{\omega_1^H \ell_1^J + \left[(1-\omega_1^H)\sigma + (\omega_1^H + (1-\omega_1^H)\sigma)\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}\right]\ell_0^J}{g(\omega_1^H + (1-\omega_1^H)\sigma) + (1-g)\sigma}, \tag{21}$$

which is again decreasing in g, equal to $p > 0$ for $g = 0$, and smaller than $p + \ell_1^J < 0$ for $g = 1$. Again, this proves that there exists an $\underline{g}$ such that, in equilibrium, L does no wrong and J_1 publishes if and only if $g \leq \underline{g}$. Furthermore, (18) is equal to (21) when substituting for $\widetilde{\mu}_2 = 1$.

Hence, when requiring $\widetilde{\mu}_2 = 1$ if $-\ell_0^L \geq b + \sigma r$, we can write more generally

$$\underline{g} = \frac{\sigma p}{-\omega_1^H \ell_1^J - \left[(1-\omega_1^H)\sigma + (\omega_1^H + (1-\omega_1^H)\sigma)\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}\right]\ell_0^J - p\omega_1^H(1-\sigma)} \tag{22}$$

Clearly, $\underline{g}$ is increasing in $\widetilde{\mu}_2$ which, according to Lemma 1, is nondecreasing in p, ℓ_0^J and ℓ_1^J. Furthermore, it can immediately be seen that the direct effects of ℓ_0^J and ℓ_1^J on $\underline{g}$ are weakly

that the equilibrium is selected where $\omega_1^H = 1$ whenever it exists and $\omega_1^H = 0$ otherwise. The above discussion has proven that this equilibrium selection rule is well-defined.

positive, depending on ω_1^H (and indeed strictly positive in the case of ℓ_0^J). The effect of p on $\underline{g}$ can be seen by taking the derivative

$$\frac{\partial \underline{g}}{\partial p} = \frac{\sigma\left[-\omega_1^H\ell_1^J - \left[(1-\omega_1^H)\sigma + (\omega_1^H + (1-\omega_1^H)\sigma)\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}\right]\ell_0^J\right]}{\left[-\omega_1^H\ell_1^J - \left[(1-\omega_1^H)\sigma + (\omega_1^H + (1-\omega_1^H)\sigma)\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}\right]\ell_0^J - p\omega_1^H(1-\sigma)\right]^2} > 0.$$

Furthermore, as $\widetilde{\mu}_2$ is determined by the requirement that L must be indifferent between suing and not suing in period 1, it also depends (step-wise) on σ, b, r and ℓ_0^L. As for the last of these parameters, higher ℓ_0^L makes suing for libel more attractive, so that a lower $\mu_2^S = \widetilde{\mu}_2$, which is associated with higher probabilities that J_2 publishes after observing that J_1 has been sued, is required to make suing for libel less attractive again. Therefore, $\underline{g}$ is weakly decreasing in ℓ_0^J.

However, type H's decision whether to do wrong in the first period impacts on $\underline{g}$, but in a way that cannot be generally pinned down but rather depends on the parameters. More specifically, $\underline{g}$ is larger when $\omega_1^H = 0$ rather than $\omega_1^H = 1$ if and only if $p(1-\sigma) < -\ell_1^J + \sigma\ell_0^J - \frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}(1-\sigma)\ell_0^J$, which may or may not be the case, depending on the parameters. Hence, rather than introducing additional assumptions, we conclude that all the comparative statics derived above are only local for given $\omega_1^H = 1$ or $\omega_1^H = 0$.

Consider now an equilibrium in which J_1 never publishes. In such an equilibrium, both types of public figure always do wrong in the first period. Whether J_1 indeed prefers not to publish in that case depends on what can be expected to happen off the equilibrium path after a first-period publication. Recall that, whenever $0 \leq \ell_1^H$, $\pi_2(0) = -\sigma r$, $\pi_2(1) = b$ and $\mu_2^N = 0$, so that $\pi_2(1) - \pi_2(\mu_2^N) = b + \sigma r$. Hence, if $-\ell_1^L < b + \sigma r$, it is possible to find some $\widetilde{\mu}_2$ such that type L wants to sue in the first period if and only if, in equilibrium, $\mu_2^S \geq \widetilde{\mu}_2$. Furthermore, as both types always do wrong in the first period and type H always sues, $\mu_2^S = \frac{g}{g+(1-g)\lambda_1^L}$, so that $\mu_2^S \geq \widetilde{\mu}_2$ if and only if $\lambda_1^L \leq \frac{g}{1-g}\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}$, the right-hand side of which is increasing in g (as $\widetilde{\mu}_2$ is independent of g). On the other hand, substituting for $\omega_1^H = \omega_1^L = 1$ in (11) implies that J_1 does not publish if $p + (g + (1-g)\lambda_1^L)\ell_1^J < 0$, which is equivalent to $\lambda_1^L > \frac{p+g\ell_1^J}{-(1-g)\ell_1^J}$, the right-hand side of which is decreasing in g. There exists a λ_1^L that satisfies both of these conditions if and only if $\frac{p+g\ell_1^J}{-(1-g)\ell_1^J} < \frac{g}{1-g}\frac{1-\widetilde{\mu}_2}{\widetilde{\mu}_2}$, which is equivalent to

$$g > \overline{g} := \frac{p}{-\ell_1^J}\widetilde{\mu}_2 \tag{23}$$

If $-\ell_1^L \geq b + \sigma r$, then, in equilibrium, $\lambda_1^L = 0$, in which case J_1 does not publish if and only if $g > \overline{g} = \frac{p}{-\ell_1^J}$. In both cases, the proposition's claim that an equilibrium in which J_1 does not publish and type L does wrong exists if and only if g is above some threshold $\overline{g}$. Again, $\overline{g}$ is strictly increasing in p, ℓ_1^J and $\widetilde{\mu}_2$ which, in turn, is nondecreasing in p, ℓ_0^J and ℓ_1^J. Just as in the above comparative-statics analysis of $\underline{g}$, larger ℓ_1^L makes suing for libel more attractive, which may require reducing $\widetilde{\mu_2}$, so that $\overline{g}$ is step-wise decreasing in ℓ_1^L.

References

Acheson, D. J., and A. Wohlschlegel (2018a): "The economics of weaponized defamation lawsuits," *Southwestern Law Review*, 47(2), 335–383.

——— (2018b): *Libel Bullies, Defamation Victims and Litigation Incentives*. Portsmouth Business School Working Paper in Economics & Finance 2018-01.

Anderson, D. A. (1975): "Libel and Press Self-Censorship," *Texas Law Review*, 53, 422–481.

Arbel, Y. A., and M. C. Mungan (2019): "The Case Against Expanding Defamation Laws," *Alabama Law Review*, 71.

Bar-Gill, O., and A. Hamdani (2003): "Optimal Liability for Libel," *Contributions in Economic Analysis & Policy*, 2(1).

Barendt, E., L. Lustgarten, K. Norrie, and H. Stephenson (1997): *Libel law and the media: the chilling effect*. Clarendon Press.

Baum, I., E. Feess, and A. Wohlschlegel (2009): "Reporter's Privilege and Incentives to Leak," *Review of Law & Economics*, 5(1), 701–715.

Bergelson, L. (2018): "The Need for a Federal Anti-SLAPP Law in Today's Digital Media Climate," *Columbia Journal of Law & the Arts*, 42, 213.

Bezanson, R. P. (1988): "The Libel Tort Today," *Washington & Lee Law Review*, 45, 535–556.

Boies, D. (1994): "The Chilling Effect of Libel Defamation Costs: The Problem and Possible Solution," *St Louis University Law Journal*, 39, 1207.

Dalvi, M., and J. F. Refalo (2008): "An Economic Analysis of Libel Law," *Eastern Economic Journal*, 34(1), 74–94.

Dent, C., and A. Kenyon (2004): "Defamation law's chilling effect: A comparative content analysis of Australian and US newspapers," *Media and Arts Law Review*, 9(89), 92.

Farmer, A., and P. Pecorino (1998): "A reputation for being a nuisance: frivolous lawsuits and fee shifting in a repeated play game," *International Review of Law and Economics*, 18(2), 147–157.

Feess, E., G. Muehlheusser, and A. Wohlschlegel (2009): "Environmental liability under uncertain causation," *European Journal of Law and Economics*, 28(2), 133–148.

Garoupa, N. (1999a): "Dishonesty and Libel Law: The Economics of the 'Chilling' Effect," *Journal of Institutional and Theoretical Economics*, 155(2), 284–300.

——— (1999b): "The Economics of Political Dishonesty and Defamation," *International Review of Law and Economics*, 19(2), 167–180.

Givati, Y. (2016): "A Theory of Whistleblower Rewards," *Journal of Legal Studies*, 45(1), 43–72.

Gratton, G. (2015): "The sound of silence: Political accountability and libel law," *European Journal of Political Economy*, 37, 266–279.

Heyes, A., and S. Kapur (2008): "An economic model of whistle-blower policy," *Journal of Law, Economics, & Organization*, 25(1), 157–182.

Hollander, D. A. (1989): "The Economics of Libel Litigation," in *The cost of libel: Economic and policy implications*, ed. by E. E. Dennis, and E. M. Noam, p. 207. Columbia University Press.

Hooper, D. (2000): *Reputations Under Fire: Winners and Losers in the Libel Business*. Little, Brown.

Hovenkamp, E. (2013): "Predatory Patent Litigation," Discussion paper.

Kenyon, A. T., and T. Marjoribanks (2008): "Chilled Journalism?: Defamation and Public Speech in US and Australian Law and Journalism," *New Zealand Sociology*, 23(2), 18.

Kim, J.-Y., and S. Lee (2019): "Apportionment of liability by the stochastic Shapley value," *International Review of Law and Economics*, 60, Article 105860.

Kim, J.-Y., and J. Song (2007): "Signal jamming in pretrial negotiation with multiple defendants," *Journal of Economics*, 91(2), 177–200.

Logan, D. A. (2001): "Libel law in the trenches: Reflections on current data on libel litigation," *Virginia Law Review*, 87, 503–529.

Lopez, A. S. (2012): "Protecting Speech, Protecting Privacy: The Future Costs of UK Libel Claims," *Loyola of Los Angeles International and Comparative Law Review*, 35, 433.

Mechtenberg, L., G. Muehlheusser, and A. Roider (2020): "Whistleblower protection: Theory and experimental evidence," *European Economic Review*, 126, 103447.

Miceli, T. J. (1993): "Optimal deterrence of nuisance suits by repeat defendants," *International Review of Law and Economics*, 13(2), 135–144.

Mullis, A., and A. Scott (2009): "Something rotten in the state of English libel law? A rejoinder to the clamour for reform of defamation," *Communications Law*, 14(6), 173–183.

Schauer, F. (1978): "Fear, Risk and the First Amendment: Unraveling the Chilling Effect," *Boston University Law Review*, 58, 685–732.

Townend, J. (2014): "Online chilling effects in England and Wales," *Internet Policy Review*, 3(2), 1–12.

——— (2019): "Unreported scandals: the power of personality and legal bluster," in *The Routledge Companion to Media and Scandal*, ed. by H. Tumber, and S. Waisbord, pp. 202–209. Routledge.

Vick, D. W., and L. Macpherson (1995): "Anglicizing Defamation Law in the European Union," *Virginia Journal of International Law*, 36, 933.

Weaver, R. L. (2012): "British Defamation Reform: An American Perspective," *Northern Ireland Legal Quarterly*, 63, 97.

Weaver, R. L., A. T. Kenyon, D. F. Partlett, and C. P. Walker (2004): "Defamation law and free speech: Reynolds v. Times Newspapers and the English media," *Vanderbilt Journal of Transnational Law*, 37, 1255.

——— (2006): *The right to speak ill: Defamation, reputation and free speech*. Carolina Academic Press.

Windon, J. (2010): "Fee Shifting in Libel Litigation: How the American Approach to Costs Allocation Inhibits the Achievement of Libel Law's Substantive Goals," *Journal of International Media & Entertainment Law*, 3, 175–194.