Málaga Economic Theory
Research Center
Working Papers

A careerist judge with two concerns

Ascensión Andina-Díaz and José A. García-Martínez

WP 2016-2
August 2016

Departamento de Teoría e Historia Económica
Facultad de Ciencias Económicas y Empresariales
Universidad de Málaga
ISSN 1989-6908

A careerist judge with two concerns*

Ascensión Andina-Díaz[†]and José A. García-Martínez[‡]

August 1, 2016

Abstract

This paper analyzes the effect of media coverage on the incentives of a careerist judge to act on her information. The novelty of our approach is to consider a judge who seeks to signal both *expertise* and (absence of) *bias*. To this, we incorporate three types of judges: the high quality or wise judge, the normal judge and the biased/lenient judge. Our results show that whether media coverage of judicial cases leads to better sentencing practices or not, deeply depend on the proportion of lenient judges in the population. In particular, we obtain that when this proportion is either too low or too high, media coverage does not affect a judge's behavior. However, her behavior if very different in one case than in the other. We also obtain that when the perceived proportion of lenient judges is neither too strong nor too weak, media coverage of judicial cases can induce a judge to act less on her information and to pass harsher sentences.

Keywords: Careerist judges; bias; transparency; media coverage; sentencing practices

JEL: D82, K40, L82

1 Introduction

It is common in the literature of law and economics to consider that judges have reputational concerns. Examples are Miceli and Cosgel (1994), Levy (2005), Iossa and Jullien (2012) or Cohen et al. (2015). There are at least three reasons for this. First, because as in any other profession, judges may care about what peers in the judiciary and the general public thinks about them. Second, because reputation is an asset of special relevance to professionals of the judiciary, who rely on their authority and the reputation of the courts to have their decisions respected and complied. Third, because despite the great heterogeneity in the operation of courts across countries (see Djankov et al. (2003) and Glaeser and Shleifer (2002)), it is common to most judicial systems to elect judges for the highest levels of the hierarchical system from lower-echelon courts (See Garoupa and Ginsburg (2011) and Rubin and Shepherd (2013)).[1]

*We thank participants at ASSET Granada 2015 for comments. We gratefully acknowledge financial support from the Ministerio de Educación y Ciencia through projects ECO2010-21624 and ECO2011-26996, and the Junta de Andalucía through project SEJ2011-8065. The usual disclaimer applies.

[†]Dpto. Teoría e Historia Económica, Universidad de Málaga, 29071 Málaga, Spain. E-mail: aandina@uma.es

[‡]Dpto. Estudios Económicos y Financieros, Universidad Miguel Hernández, Elche, Alicante, Spain. E-mail: jose.garciam@umh.es

[1]In general, judges for appellate courts are usually promoted from trial courts, and judges for the courts of last report or supreme courts are usually appointed from appellate courts. This is generally the case in countries such as UK or US (with a common law system), and France or Spain (with a civil law system).

Previous research on career concerns show that when an agent (in this case a judge) seeks to maximize her reputation for being high quality, increasing the probability that the principal (in this case, the voters, the governor or any other judge's evaluator) receives information on the correct action, increases the probability that the agent follows her informative signals and makes more accurate decisions. Accordingly, at least since Prat (2005), we say that transparency of consequences -giving the evaluator information on the correct action- is beneficial.

Because citizens receive most of their information from the media, the previous argument suggests that we may expect more accurate sentencing practices in court cases receiving media coverage. This is in line with Lim (2015), who observes that for US state courts, differences in damage awards between courts in conservative and liberal districts reduce with media coverage. However, empirical evidence shows that media coverage can also produce perverse effects. In fact, a recent study by Lim et al. (2015) for US state court judges, shows that newspaper coverage significantly increases sentence length by nonpartisan elected judges, and that this effect is increasing in the severity of crimes. They quantify the effect and find that for the period $1986-2006$, eight more newspaper articles per judge per year in a judicial district increases average sentence length for violent crimes by nonpartisan elected judges by 3.4%, which means a sentence of about 5.7 months longer.[2]

Outside the United States, the thesis that media coverage of courts may distort judges' decisions and induce them to pass harsher sentences, finds its support in the following examples. The first one refers to "*Caso Pantoja*". In 2007, Isabel Pantoja, an internationally renowned Spanish *copla* singer, was accused of laundering money proceeding from public funds profited by her partner during Spain's property boom. In April 2013, she was sentenced by the provincial court of Málaga, Spain, to two-year in prison. Though initially she was not expected to go to jail (no prior convictions and minimum prison-sentence, in which case most defendants avoid jail), the court rejected all her appeals to have the sentence suspended and she finally entered prison in January 2015. In the sentence, the court explicitly said "*it was making an example of Pantoja*" at a time of economic crisis and rampant corruption by Spanish public figures.[3]

The second example is the "*Nut rage*" trial. The facts took place in December 2014, at John F. Kennedy Airport in NY, when Cho Hyun-ah, executive of Korean Air and daughter of the company's chief, required the plane to return to gate before takeoff, after she was served nuts in a bag. The apparently minor air rage sparked public outrage in South Korea, where "*the incident was seen as emblematic of a generation of spoilt and arrogant offsprint of owners of family-run conglomerates that dominate the economy*".[4] Two months later, Cho Hyun-ah was sentenced by the Seoul Western District Court, South Korea, to one-year prison for obstructing aviation safety.

In this paper we investigate which incentives can lead a careerist judge to deviate more often from her interpretation of the law and to pass harsher sentences in cases that receive media coverage than in cases with no resonance beyond the courtroom. To this, we consider a judge that seeks to maximize her reputation for expertise or high quality. This is as usual. The novelty of our approach is that we add a second concern: not to be perceived as biased. This assumption is based on the regularity that citizens all over the world tend to consider justice as too lenient (biased) and on the conjecture that the general

[2] The effect is significant after controlling for: (i) more severe crimes attract more media attention, and (ii) more populated districts with more newspapers might have different crime rates. In contrast, there is no significant effect of media coverage on judges competing in partisan elections, nor for appointed judges.

[3] "Una sentencia ejemplarizante", *El País*, November 8th 2014.

[4] " 'Nut rage' trial: Korean Air executive treated crew 'like slaves' ", *The Guardian*, 2nd February 2015.

public perception of a lenient justice may affect a judge's sentencing practices. With respect to the first point, worldwide opinion polls support our claim.[5] With respect the second, a numerous group of papers show that US elected judges tend to pass more punitive sentences as standing for reelection approaches (see Gordon and Huber (2007); Huber and Gordon (2004), Caldarone et al. (2009), Canes-Wrone et al. (2012) and Park (2014)). We interpret this result as clear evidence that public opinion can condition sentencing practices. Interestingly, we show that the interaction between these two objectives, expertise and absence of bias, explains why increasing transparency is not necessarily always good.

The model has the following structure. A judge (agent) has to sentence a convicted offender. There are two possible sentences: a light or lenient sentence, and a harsh or heavy sentence. Accordingly, there are two possible states of the world: one in which the correct sentence is light, the other one in which the severe sentence corresponds. The judge receives a signal about the state, with a quality that depends on the ability of the judge to interpret the law, and imposes a punishment. Alternatively, the quality of a signal can also be interpreted as the difficulty of a particular case, which may vary depending on available evidence. Trying to illustrate the trade-off for a judge between passing the correct sentence (showing ability) and avoiding being perceived as too lenient (or biased), we consider a judge with a double concern: a concern for expertise, and a concern for bias. To this, the model incorporates three types of judges. A high quality or wise judge, who receives a perfectly informative signal that always follows; a normal judge, whose signal is informative though imperfect; and a biased judge, who is assumed to have a preferred sentence (light punishment) she always passes. The principal (could be the peers in the judicature, a governor, the general public, or any other audience that could evaluate the judge's performance) observes the punishment and updates his belief about the type of the judge, based on the sentence and, if available, on the ex-post verification of the state.

With this, we want to illustrate the idea that whereas for certain judicial cases the principal does not receive any feedback about whether the imposed punishment is fair or not, for other situations the principal may learn whether, according to law, the sentence is correct. In this case we say there is full transparency of consequences. In the real world, transparency of consequences occurs when a sentence is appealed to a higher judicial body that renders a decision, or when a case receives the attention of the media (for example because of the name of the defendant or the type of the crime), and the details of the case and the appropriate punishment is discussed by experts and covered by the media. In contrast, when a case has no echo beyond the courtroom, it seems unlikely that peers in the judiciary and citizens receive any new information to confront with the observed punishment. For expositional purposes, in the paper we will identify the principal with the general public and will assume that media coverage increases the probability that the general public learns the correct sentence. Then, note that in the present paper the probability that the state is publicly observed is exogenous to the judge. This is in contrast to Levy (2005), who considers that a judge can affect the probability that a sentence is brought before a higher court and thus, decide on the probability that the evaluator receives ex-post verification of the state.[6]

The objective of the paper is twofold. First, to characterize the behavior of a judge who cares both, for expertise and bias, and to investigate whether her behavior varies with the general public perception

[5]For example, the study conducted by Princeton Survery Reseach Associates International in 2006 for the US National Center for State Courts finds that: "*More Americans are inclined to say sentencing practices in their state generally are too lenient than believe they are too harsh (*48% *vs.* 8%*)*". Evidence for Australia, Canada, Germany, Spain or UK shows similar patterns.

[6]See also Andina-Díaz and García-Martínez (2015), who considers endogenous monitoring in the media industry.

of a lenient justice. We obtain that the judge's behavior varies indeed with the perceived proportion of lenient judges in the judicial system. In particular, we obtain that when the perception of a lenient justice is low enough, in equilibrium the judge always follow her signal, to show her expertise. In contrast, when the perception of a lenient justice is too high, in equilibrium the judge always passes harsh sentences, for fear of being tarred as soft on crime. Last, when the public perception of a lenient justice is neither too high nor too low, we obtain that the judge's behavior is more complex, as it also depends on whether the general public will learn the correct sentence or not.

Second, to analyze the effects of media coverage -that increases the probability that the general public learns about the correct sentence- on the judge's sentencing practices. We obtain that if judges have an only concern (either for expertise or for bias), media coverage never results in less accurate judges' practices. As for the case in which judges have two concerns, our results show that whether media coverage leads to better sentencing practices or not, deeply depends on the perceived proportion of lenient judges in the judicial system. In particular, we obtain that when this proportion is either too low or too high, media coverage of judicial cases does not affect judges' sentencing practices. In contrast, when there is a moderate perception of a lenient justice and the judge is not specially able to interpret the law or, alternatively, the case is specially difficult (for example, because of lack of conclusive evidence), our results show that media coverage of judicial cases can induce judges to act less on their information and more precisely, to pass harsher sentences.

We consider that the result that media coverage may induce judges to deviate from their informative signals and to pass less accurate sentences makes a contribution to both, the literature on law and economics and the literature on the effects of transparency on careerist agents. To the former, our model proposes a theory to explain the empirical findings in Lim et al. (2015), that is why newspaper coverage can increase sentence length. A result that finds additional support in the aforementioned examples of media publicized cases receiving specially harsh sentences. To the latter, our model proposes a new argument to add to the already existing in the literature (see Prat (2005), Bourjade and Jullien (2011), Fox and Van Weelden (2012) or Andina-Díaz and García-Martínez (2015) among others) that explains why transparency can have perverse effects.[7] The novelty is to consider an expert with two concerns and to identify where and why transparency can be here detrimental in terms of social welfare. To the best of our knowledge, this is a novel result in the literature.

The rest of the paper is organized as follows. The next section describes the model. In Section 3 we analyze the behavior of the judge when she has an only concern. The analysis of these scenarios will serve us as benchmarks for the posterior analysis. In Section 4 we present the results for the case where the judge has two concerns. Section 5 concludes. All the proofs are relegated to Appendix A and Appendix B analyzes the model with a strategic high type judge.

[7] Prat (2005) shows that transparency of actions can be detrimental to the principal. Bourjade and Jullien (2011) show that when an expert is biased, increasing transparency on the preferences of the expert does not induce her to reveal more information. Fox and Van Weelden (2012) show that when the prior of the state is sufficiently strong, transparency of consequences can make an expert more reticent to act on her information. Last, Andina-Díaz and García-Martínez (2015) show that when the probability of transparency is endogenous to the expert's action, increasing the probability that the principal learns the state induces the expert to act less on her information.

2 The model

A judge (agent) has to sentence a convicted offender. There are two possible sentences or punishments: a light or lenient sentence, denoted by $\hat{l}$, and a heavy or harsh sentence, denoted by $\hat{h}$. Accordingly, there are two states of the world, $\omega \in \{L, H\}$. We assume the lenient sentence $\hat{l}$ is the correct punishment in state L, and the harsh sentence $\hat{h}$ is the appropriate one in state H. Each state occurs with equal probability.[8]

The judge receives a signal $s \in \{l, h\}$ about the correct interpretation of the law, and passes a sentence. The quality of the signal depends on the judge's ability to interpret the law. Let $\gamma = P(l/L) = P(h/H) > \frac{1}{2}$ be the judge's ability. Alternatively, γ can be interpreted as the degree of difficulty of the case. Under this interpretation, the smaller the value of γ, the less conclusive evidence is and so, the more difficult the case is. We assume that the judge observes the signal $s \in \{l, h\}$ and takes an action $a \in \{\hat{l}, \hat{h}\}$. We refer to the action as the judge's sentence.

At the time to pass the sentence, the type of the judge is her private information. Judges can be of high expertise, in which case the judge gathers a perfectly informative signal that always follows. We will refer to this type as the wise judge, and will denote it by W.[9] Judges can also be lenient or soft on crime, in which case the judge always passes a light sentence, independently of her signal.[10] We will refer to this type as the biased judge, and will denote it by B. Last, judges can be normal, in which case the judge has career concerns and seeks both, to maximize the probability of passing the correct sentence (a concern for *expertise*) and to minimize the probability of being perceived as a lenient judge soft on crime (a concern for *bias*). We will denote by N the normal type. Let α_W, α_B and α_N denote the prior probability that the general public assigns to a judge being a wise, a biased or a normal type, respectively, with $\alpha_W + \alpha_B + \alpha_N = 1$.

The focus of the paper is on the behavior of normal judges, hereafter simply referred as the judge. We denote by $\sigma_s(a) \in [0, 1]$ the probability that, conditioned on her signal $s \in \{l, h\}$, the judge takes action $a \in \{\hat{l}, \hat{h}\}$. Note that action $\hat{h}$ is never taken by the biased type. In this sense, in the paper we will refer to action $\hat{h}$, the harsh sentence, as the *politically correct* action; and to action $\hat{l}$, the lenient sentence, as the *politically incorrect* action.

The principal or evaluator represents those that the judge would like to impress. In our case, we will identify it with the general public. Alternatively, it could be a governor, the peers in the judiciary or any other player whose opinion is important to the judge.

As usual in career concerns models, we assume that the principal observes the sentence $a \in \{\hat{l}, \hat{h}\}$ and forms a belief about the type of the judge. Additionally, we consider that prior to forming the belief, there is a probability $\mu \in [0, 1]$ that the principal gleans information about the true state of the world. Note that if $\mu = 1$ we are in a situation in which the general public learns the correct sentence with probability 1, whereas $\mu = 0$ represents a situation in which the state ω is not revealed. Naturally, the

[8] See the discussion in the Conclusion.

[9] In Appendix B we extend the analysis and consider the case in which the wise judge is also strategic, as the normal judge. We obtain that there is an equilibrium in which the wise type follows her signal and passes the correct sentence, conditioned on the normal type does not always take action $\hat{h}$, independently of her signal. In this case, the equilibrium is also unique. In the less interesting scenario in which the normal type always takes action $\hat{h}$, we show that the honest strategy for the wise judge is an equilibrium strategy if α_B is below a certain threshold.

[10] See the discussion in the Conclusion.

information on the correct sentence is taken into account by the general public when assessing a belief about the judge's expertise and bias. Following Prat (2005), we will sometimes refer to probability μ as the level of transparency (of consequences).

For expositional purposes, in the paper we will assume that media coverage increases the probability μ that the general public learns the correct sentence. Two interpretations are here permitted. Under the first interpretation, μ is simply the probability that a judicial case receives media coverage. This interpretation accounts for situations in which the judge, at the time to pass a sentence, may not know whether the sentence will attract media coverage and so, whether the general public will count with additional information to asses the judge's validity. Note that this interpretation implicitly assumes that citizens can only learn about the correct sentence from the media. In contrast to this, it could be the case that citizens have other sources of information (for example their own reading of a sentence) and that media coverage of judicial cases simply increases the probability that the general public learns the state. This is the second interpretation that our model permits. Note that under both interpretations the effect of media coverage of judicial cases is always to increase the probability that citizens learn the correct sentence. This is all that matters for our discussion.

We denote by $X \in \{0, L, H\}$ the information gleaned by citizens about the state of the world, where L (respectively H) means that the general public learns that the correct sentence is $\hat{l}$ (respectively $\hat{h}$); and 0 means that the public learns nothing.

The general public observes sentence $a \in \{\hat{l}, \hat{h}\}$ and feedback $X \in \{0, L, H\}$ and, based on this information, update their belief about the type of the judge. Let $\lambda_W(a, X)$ denote the principal's posterior probability that the judge is a wise or high quality type, and let $\lambda_B(a, X)$ denote the posterior probability that she is a lenient-biased type. According to Bayes' rule:

$\lambda_W(a, X)$			
	0	L	H
$\hat{l}$	$\frac{\alpha_W}{2\alpha_B+\alpha_W+(\sigma_h(\hat{l})+\sigma_l(\hat{l}))\alpha_N}$	$\frac{\alpha_W}{\alpha_B+\alpha_W+((1-\gamma)\sigma_h(\hat{l})+\gamma\sigma_l(l))\alpha_N}$	0
$\hat{h}$	$\frac{\alpha_W}{\alpha_W+(\sigma_h(\hat{h})+\sigma_l(\hat{h}))\alpha_N}$	0	$\frac{\alpha_W}{\alpha_W+(\gamma\sigma_h(\hat{h})+(1-\gamma)\sigma_l(\hat{h}))\alpha_N}$

Table 1

$\lambda_B(a, X)$			
	0	L	H
$\hat{l}$	$\frac{2\alpha_B}{2\alpha_B+\alpha_W+(\sigma_h(\hat{l})+\sigma_l(\hat{l}))\alpha_N}$	$\frac{\alpha_B}{\alpha_B+\alpha_W+((1-\gamma)\sigma_h(\hat{l})+\gamma\sigma_l(\hat{l}))\alpha_N}$	$\frac{\alpha_B}{\alpha_B+(\gamma\sigma_h(\hat{l})+(1-\gamma)\sigma_l(\hat{l}))\alpha_N}$
$\hat{h}$	0	0	0

Table 2

As already discuss, we assume a careerist judge that seeks both, to maximize the probability that the principal places on her being a wise judge, and to minimize the probability of being perceived as a lenient judge soft on crime. Her objective function is assumed to be:

$$f(\lambda_W(a, X), 1 - \lambda_B(a, X)) = f\left(\lambda_W(a, X), \lambda_{\bar{B}}(a, X)\right), \tag{1}$$

with $\lambda_{\bar{B}}(a,X) = 1 - \lambda_B(a,X)$ being the probability of not being considered a biased type.

The general public is concerned with justice or social welfare, and receives utility 1 when the convicted offender receives the appropriate sentence, that is $a = \omega$; and 0 otherwise. Note that because the judge's signal is informative, i.e. $\gamma > 1/2$, in the eyes of the general public and the society as a whole, the best the judge can do is to make an appropriate interpretation of the law and to pass the sentence that her signal indicates.

Prior to the analysis of the game, we introduce some definitions. Given the probability $\mu \in [0,1]$ that the state is publicly observed, let $\Pi_s^\mu(a)$ be the expected payoff to the judge when she receives signal $s \in \{l,h\}$ and passes sentence $a \in \{\hat{l},\hat{h}\}$. Thus,

$$\Pi_s^\mu(a) = (1-\mu)\Pi_s^0(a) + \mu\Pi_s^1(a), \tag{2}$$

where $\Pi_s^0(a)$ denotes the payoff to a judge that receives signal s and takes action a when the state ω is not publicly observed ($\mu = 0$), and $\Pi_s^1(a)$ denotes her (expected) payoff when the state is publicly observed ($\mu = 1$). Using equation (1) we have:

$$\Pi_s^0(a) = f\left(\lambda_W(a,0), \lambda_{\bar{B}}(a,0)\right), \tag{3}$$

$$\Pi_l^1(a) = \gamma f(\lambda_W(a,L), \lambda_{\bar{B}}(a,L)) + (1-\gamma) f(\lambda_W(a,H), \lambda_{\bar{B}}(a,H)), \tag{4}$$

$$\Pi_h^1(a) = (1-\gamma) f(\lambda_W(a,L), \lambda_{\bar{B}}(a,L)) + \gamma f(\lambda_W(a,H), \lambda_{\bar{B}}(a,H)), \tag{5}$$

Additionally, let us define the expected gain to passing sentence $\hat{h}$ rather than $\hat{l}$, after observing signal s, when the probability of having ex-post verification of the state is $\mu \in [0,1]$, as:

$$\Delta_s^\mu = \Pi_s^\mu(\hat{h}) - \Pi_s^\mu(\hat{l}), \tag{6}$$

which can be rewritten as:

$$\Delta_s^\mu = (1-\mu)\Delta_s^0 + \mu\Delta_s^1 \tag{7}$$

with $\Delta_s^0 = \Pi_s^0(\hat{h}) - \Pi_s^0(\hat{l})$ and $\Delta_s^1 = \Pi_s^1(\hat{h}) - \Pi_s^1(\hat{l})$.[11]

Our equilibrium concept is perfect Bayesian equilibrium. In the following, we will say that $(\sigma_l^\mu(\hat{l}), \sigma_h^\mu(\hat{h}))$ is an *equilibrium strategy* if $\sigma_l^\mu(\hat{l})$ maximizes the expected payoff to the judge after observing signal l, and $\sigma_h^\mu(\hat{h})$ does it after signal h, when the level of transparency is $\mu \in [0,1]$. We will denote an equilibrium strategy by $\sigma^{\mu*} = (\sigma_l^\mu(\hat{l})^*, \sigma_h^\mu(\hat{h})^*)$.

Additionally, we will say that an equilibrium is *informative* if $\sigma_l^\mu(a)^* \neq \sigma_h^\mu(a)^*$, for any $a \in \{\hat{l},\hat{h}\}$ and $\mu \in [0,1]$, and *non-informative* otherwise. An informative equilibrium is *non-perverse* if $\sigma_l^\mu(\hat{l})^* > \sigma_h^\mu(\hat{l})^*$ and $\sigma_h^\mu(\hat{h})^* > \sigma_l^\mu(\hat{h})^*$. We focus on non-perverse equilibria.

3 Benchmark analysis

The objective of this paper is twofold. First, to characterize the behavior of the judge as a function of the general public perception of the proportion of lenient judges soft on crime in the judicial system. Second,

[11]Note that Δ_s^μ is a function of the strategy profile $(\sigma_l^\mu(\hat{l}), \sigma_h^\mu(\hat{h}))$. For simplicity, this dependence is sometimes omitted.

to show that media coverage of judicial cases can induce a judge to act less on her signal and so, to pass less accurate sentences.

Before solving for the equilibria of the model and showing this perverse effect of transparency, it is useful to investigate how a judge with a sole concern, either for expertise or for bias, would take her decisions, and the role of media coverage in this case. This study will serve as benchmarks for the posterior analysis.

To facilitate comparisons with Section 4, all the other aspects of the model remain the same. Specifically, we consider the existence of three types of judges.[12]

■ **A concern for bias:** Let us start considering the scenario in which the judge only seeks to minimize the probability of being perceived as a lenient type soft on crime. From equations (2) − (5), we can rewrite the expected payoff to a judge for passing action $a \in \{\hat{l}, \hat{h}\}$ after signal $s \in \{l, h\}$, when the level of transparency is μ, as:

$$\Pi_s^\mu(a) = (1-\mu)\lambda_{\bar{B}}(a, 0) + \mu(P(L \mid s)\lambda_{\bar{B}}(a, L) + P(H \mid s)\lambda_{\bar{B}}(a, H)).$$

The analysis of this case is quite simple. In fact, note that because action h is never taken by the biased type, the payoff to a judge that passes the harsh sentence is always 1, independently of signal s and probability μ. In contrast, a judge that passes the lenient sentence can not avoid being missed with the lenient type, even if the state is shown to be L and the general public learns it. Hence, the judge's payoff is smaller than 1 in this case. Accordingly, a judge with an only concern for bias will always pass the politically correct sentence $\hat{h}$. That is, for any $\mu \in [0, 1]$, the unique equilibrium strategy $(\sigma_l^\mu(\hat{l})^*, \sigma_h^\mu(\hat{h})^*)$ of a bias-concerned judge is $(0, 1)$.

Note that in this case the judge's behavior does not depend on the probability μ that the general public learns the state. Hence, we can also conclude that in a model in which the judge has an only concern for bias, media coverage has no perverse effect on sentencing practices.

■ **A concern for expertise:** Let us now consider the case where the judge has an only concern for showing her expertise or ability as a type that correctly interprets the law. In this case, the expected payoff to the judge for passing action $a \in \{\hat{l}, \hat{h}\}$ after signal $s \in \{l, h\}$, when the level of transparency is μ, is:

$$\Pi_s^\mu(a) = (1-\mu)\lambda_W(a, 0) + \mu(P(L \mid s)\lambda_W(a, L) + P(H \mid s)\lambda_W(a, H)).$$

The analysis of this case is a bit more complex. For this reason, we first analyze the behavior of the judge when there is no transparency, i.e. $\mu = 0$, then we consider the case with full transparency, i.e. $\mu = 1$ and last, we analyze the equilibrium behavior when $\mu \in (0, 1)$. We obtain the following results:

Lemma 1. *Suppose* $\mu = 0$.

(i) If $\alpha_B \geq \alpha_N$*, there is a unique equilibrium. In the equilibrium,* $\sigma_l^0(\hat{l})^* = 0$ *and* $\sigma_h^0(\hat{h})^* = 1$.

(ii) If $\alpha_B < \alpha_N$*, there is a unique equilibrium. In the equilibrium* $\sigma_l^0(\hat{l})^* = 1 - \frac{\alpha_B}{\alpha_N}$ *and* $\sigma_h^0(\hat{h})^* = 1$

[12] As for equation (1), we assume $f(\cdot)$ to be $\lambda_{\bar{B}}(a, X)$ in the first case and $\lambda_W(a, X)$ in the second case.

We observe that if the judge has an only concern for expertise and the state is never publicly observed, she always passes the harsh sentence when her signal indicates h.[13] However, we observe that she does not always follow a lenient signal and, in the limit, when the prior on the proportion of biased types in the population is sufficiently high, she only passes harsh sentences. This is an interesting result. In fact, it says that even if a judge does not care at all about being perceived as soft on crime, the general public perception that justice is lenient will induce her to deviate from the politically incorrect sentence and to pass harsher sentences more often than she should.

Next, we consider the case where the state ω is always publicly observed, i.e. $\mu = 1$.

Lemma 2. *Suppose $\mu = 1$. There is a unique equilibrium. In the equilibrium, $\sigma_h^1(\hat{h})^* = 1$ and either:*

$$\sigma_l^1(\hat{l})^* = \begin{cases} 1 & if \quad \alpha_B \leq \bar{\alpha}_B^0 \\ 0 < \frac{\gamma(\alpha_W+\alpha_N)+(\gamma-1)(\alpha_B+\alpha_W)}{2\gamma\alpha_N(1-\gamma)} < 1 & if \quad \bar{\alpha}_B^0 < \alpha_B < \bar{\alpha}_B^1 \\ 0 & if \quad \alpha_B > \bar{\alpha}_B^1 \end{cases}$$

with $\bar{\alpha}_B^0 = \frac{(2\gamma-1)(\gamma(1-\alpha_W)+\alpha_W)}{1-(1-\gamma)2\gamma}$, $\bar{\alpha}_B^1 = \gamma - (1-\gamma)\,\alpha_W$ *and* $\bar{\alpha}_B^0 < \bar{\alpha}_B^1$.

We observe that also with full transparency, a judge that receives signal h always passes the harsh sentence. In this sense, giving the general public information on the correct sentence, for example through the media, seems not to affect the judge's behavior after signal h. This result will be shown to be more general next. On the other hand, Lemma 2 also shows after signal l the judge's behavior deeply depend on the prior probability that a judge is biased. Thus, if the general public perceives that a high proportion of judges are lenient, a careerist judge will also here always deviate from the politically incorrect sentence. Hence, an increase in transparency does not affect the judge's behavior in this case. In contrast, if the general public perception is that only a small proportion of judges are soft on crime, a careerist judge will now find it optimal to stick to her signal with positive probability (for α_B small enough, she will always follow signal l). Here, increasing the probability that citizens learn the correct sentence seems to help increase the (a priori) accuracy of sentencing practices. Next result shows it is indeed the case.

Lemma 3. *Suppose $\mu \in (0,1)$. There is a unique equilibrium. In the equilibrium, $\sigma_h^\mu(\hat{h})^* = 1$ and $\sigma_l^\mu(\hat{l})^* \in [0,1]$. Additionally, when $\sigma_l^\mu(\hat{l})^* \in (0,1)$, then $\frac{d\sigma_l^\mu(\hat{l})^*}{d\mu} > 0$.*

To summarize, the present analysis reveals that when the judge has an only concern for expertise, media coverage of judicial cases -that increases the probability that the general public learns the correct sentence- always induces the judge to stick more often to her signal and so, to pass (a priori) more accurate sentences.

Now, taking together the analysis of two benchmark scenarios, we can conclude that media coverage of sentencing practices has never here a perverse effect, namely it is neither detrimental to the general public not to the offender. Thus, if media coverage is next to have a perverse effect, it will be because of the interplay between the two concerns. We see it next.

[13] The fact that the judge always takes action $\hat{h}$ after signal h is more general and holds for any function f, as defined by equation (1), that is increasing in both arguments. That is, to any function $f(x,y)$ such that either $f'_x(\cdot) > 0$ and $f'_y(\cdot) \geq 0$ or $f'_x(\cdot) \geq 0$ and $f'_y(\cdot) > 0$. See Proposition 1 and Remark 1 in Section 4. In the proof of the present result we use Remark 1.

4 Analysis

Let us now consider the case where the judge has both, a concern for expertise and a concern for bias. Our first result here refers to the behavior of the judge after signal h. It says that also in this case, when there is a positive probability that the general public learns the state, the judge always passes harsh sentences after signal h.[14]

Proposition 1. *Consider $\mu \in (0,1]$. For any increasing function $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X))$, in equilibrium $\sigma_h^{\mu}(\hat{h})^* = 1$.*

Proposition 1 shows that for any positive level of transparency μ and judge's ability γ, a judge with two concerns always passes the politically correct sentence when her signal indicates h. That is, as long as there is a positive probability that the general public learns the state ω, if the signal that a judge receives indicates that the convicted offender must receive a harsh punishment, she always passes harsh sentences. The intuition for this result is clear. On the one hand, taking the politically correct action guarantees the judge that the general public will never perceive her as a lenient agent. Additionally, taking action $\hat{h}$ after signal h maximizes the probability of passing the correct sentence and so the probability of being perceived as a wise judge.

Now, what about the case without transparency? Note that with $\mu = 0$ the general public can never learn the correct sentence. Hence, because no consequences are observed, the judge's expected gain to passing sentence $\hat{h}$ rather than $\hat{l}$ is the same independently of the signal. That is $\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \Delta_h^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})]$. Mathematically, we have in this case two variables and just one equation, which implies that there are multiplicity of equilibria, one in which $\sigma_h^0(\hat{h})^* = 1$. Now, because from Proposition 1 we know that any small perturbation in μ drives the judge to pass the harsh sentence, i.e. $\sigma_h^{\mu}(\hat{h})^* = 1$, we can select $\sigma_h^0(\hat{h}) = 1$ as the most robust strategy when $\mu = 0$. Accordingly, in the paper we consider that for any $\mu \in [0,1]$, $\sigma_h^{\mu}(\hat{h})^* = 1$.

Remark 1. *For any $\mu \in [0,1]$, $\sigma_h^{\mu}(\hat{h})^* = 1$.*

The result in Remark 1 is of special relevance for the posterior analysis. In fact, it says that when the judge receives signal h she always sticks to her private information and passes the harsh sentence, independently of the probability that the general public learns the correct sentence. Hence, media coverage can never have here a perverse effect.[15] The implication is straightforward. If we are interested in identifying where and why media coverage of sentencing practices can have a perverse effect, we have to focus our attention on the behavior of the judge when receiving a signal that calls for the lenient sentence. Accordingly, the rest of the paper analyzes the judge's behavior in this information set.

Suppose therefore that the judge receives signal l. Now, what sentence should the judge pass? Does her sentencing practices depend on whether the case receives media coverage or not? Could it be that

[14] For expositional purposes, in the text we first present the analysis of the benchmark scenarios and then introduce the result in Proposition 1 and the posterior Remark 1. However, in the Appendix we first characterize the behavior of a judge -with an objective function $f(\cdot)$ as defined in equation (1) that is increasing in both argument- after signal h (proof of Proposition 1) and then use this result and Remark 1 to prove Lemmas (1)-(3).

[15] Note that even if we consider the multiplicity of equilibria that we have when $\mu = 0$, transparency of consequences is never bad, since for $\mu \in (0,1]$, $\sigma_h^{\mu}(\hat{h})^* = 1$.

media coverage of judicial cases induces judges to pass (a prior) less accurate sentences? As shown next, the answer to these questions depends on the prior probability that a judge is biased or, to say it differently, on how extended the general public perception of a lenient justice is.

■ **High perception of a lenient justice:** Let us first consider the case in which the general public perceives that the proportion of lenient judges in the judicial system is very high. Our result in this case shows that if the prior probability that a judge is bias is sufficiently high, specifically $\alpha_B > \overset{\sup}{\alpha}_B$, the judge always chooses to deviate from signal l and passes the politically correct action h for any level of transparency $\mu > 0$.

Proposition 2. *For any strictly increasing function* $f(\lambda_W(a, X), \lambda_{\bar{B}}(a, X))$ *and for any* $\bar{\mu} \in (0, 1]$*, there exists* $\overset{\sup}{\alpha}_B < 1$ *such that for any* $\alpha_B > \overset{\sup}{\alpha}_B$*, in equilibrium* $\sigma_l^\mu(\hat{l})^* = 0$ *for all* $\mu \geq \bar{\mu}$*.*

That is, in this case we will have a non-informative equilibrium in which the judge always passes harsh sentences, independently of her signal, in an attempt not to be missed with the lenient judge (so common in the system). The equilibrium strategy $(\sigma_l^\mu(\hat{l})^*, \sigma_h^\mu(\hat{h})^*)$ is thus $(0, 1)$ for all μ.

Note that from Proposition 2 we also learn that media coverage of judicial cases does not here affect the judge's sentencing practices and so, it has neither beneficial nor detrimental effects to the general public.

■ **Low perception of a lenient justice:** Let us now consider the case in which the general public perceives that the proportion of lenient judges in the judicial system is very low. Our result in this case says that if the prior probability that a judge is bias is sufficiently low, specifically $\alpha_B < \overset{\inf}{\alpha}_B$, and the ability of the judge is sufficiently high, specifically $\gamma > \overset{\sup}{\gamma}$, the judge always chooses to pass the lenient sentence after signal l, for any level of transparency μ.

Proposition 3. *For any strictly increasing function* $f(\lambda_W(a, X), \lambda_{\bar{B}}(a, X))$ *and for any* $\bar{\mu} \in (0, 1]$*, there exists* $\overset{\inf}{\alpha}_B > 0$ *and* $\overset{\sup}{\gamma} < 1$ *such that for any* $\alpha_B < \overset{\inf}{\alpha}_B$ *and* $\gamma > \overset{\sup}{\gamma}$*, in equilibrium* $\sigma_l^\mu(\hat{l})^* = 1$ *for all* $\mu > \bar{\mu}$*.*

That is, we are here in a situation in which the proportion of biased types is so small and the judge's ability so high (or, alternatively, the interpretation of the law is so clear in the case) that the existence of lenient judges in the population is not sufficient to distort the incentives of a careerist judge to show her competence. Thus, we have here an informative equilibrium in which the judge always follows her signal and passes the (a priori) more accurate sentence. The equilibrium strategy $(\sigma_l^\mu(\hat{l})^*, \sigma_h^\mu(\hat{h})^*)$ is thus $(1, 1)$ for all μ.

Note that from Proposition 3 we also learn that media coverage of judicial cases does neither affect the judge's sentencing practices in this case; hence it has neither beneficial nor detrimental effects to the general public.

At this point it is clear that if transparency is to have a perverse effect, it has to be that the general public perception about a lenient justice is neither too high nor too low. That is, that for media coverage to be potentially bad, there must be a moderate perception of a biased justice in the population. Next result shows that there is a range of parameter values of α_B and γ for which this is precisely the case.

In order to obtain close-form expression, the analysis that follows assumes that the judge's objective function, as defined by equation (1), is the following:[16]

$$f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X)) = \lambda_W(a,X) + \lambda_{\bar{B}}(a,X). \tag{8}$$

The rest of the analysis is structured as follows. First, we consider a simplification of the model where either $\mu = 0$ or $\mu = 1$. This stylized scenario will allow us to build an intuition on why media coverage of judicial cases can be detrimental to the society. Then, we relax this assumption and consider the more general case $0 \leq \mu \leq 1$. We observe that the results in the first scenario maintain when we move to the second more general case.

4.1 A simple case

Let us start considering a stylized scenario in which the state ω is either publicly observed with probability one or not observed at all. The results in this section show that there are parameter configurations for which a judge that receives signal l sticks more often to her signal when the general public does not learn the correct sentence than when they do learn it.[17] Thresholds α_B^*, α_B^{**} and $\tilde{\gamma}$ are defined in the Appendix.

Proposition 4. *Let $\alpha_B^* < \alpha_B < \alpha_B^{**}$ and $\frac{1}{2} < \gamma < \tilde{\gamma}$. There is a unique equilibrium. In the equilibrium:*

- *Suppose $\mu = 0$, then $\sigma_l^0(\hat{l})^* = 1 - \frac{\alpha_B(1-\alpha_B+\alpha_W)}{(\alpha_W-\alpha_B)(1-\alpha_W-\alpha_B)} > 0$.*
- *Suppose $\mu = 1$, then $\sigma_l^1(\hat{l})^* = 0$.*

That is, from Proposition 4 we learn that when the general public perception about the proportion of lenient judges is neither too high nor too low, and the judge is not specially able to interpret the law or, alternatively, the case is specially difficult; a careerist judge will pass harsher sentences in cases receiving media coverage than in cases without the attention of the media. In other words, media coverage of judicial cases has here a perverse effect that induces judges to deviate more often from their signals and to impose more severe sentences. This is in line with the empirical findings in Lim et al. (2015) that newspaper coverage significantly increase sentence length, and with the casual observation discussed in the introduction of media publicized cases receiving specially harsh sentences.

We next try to build an intuition for this puzzling result. To this, first note that if transparency induces a judge to act less on her information, it has to be that either (i) the payoff to the judge from deviating from signal l is greater with transparency than without transparency, i.e. $\Pi_l^0(\hat{h}) < \Pi_l^1(\hat{h})$; or (ii) her payoff from following signal l is greater without transparency than with transparency, i.e. $\Pi_l^0(\hat{l}) > \Pi_l^1(\hat{l})$. In a nutshell, for our result to hold we require transparency to either increase the judge's payoff from being untruthful or decrease her payoff from being truthful.

An investigation of these two mechanisms reveals that it is not effect (i) that applies. To see it, note that for a judge with an only concern for expertise we have $\Pi_l^0(\hat{h}) > \Pi_l^1(\hat{h})$ and that for a judge with

[16] Remark 2, at the end of Appendix A, proposes a micro-foundation for the linear function.

[17] Lemma 4 in the Appendix shows that when $\mu = 0$, any judge with an increasing objective function $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X))$, does not always follow signal l. More precisely, we show that in equilibrium $\sigma_l^0(\hat{l})^* < 1$. This result extends that in Proposition 4, case $\mu = 0$, to the more general class of increasing functions f.

an only concern for bias we have $\Pi_l^0(\hat{h}) = \Pi_l^1(\hat{h})$. Hence, if we now consider a judge with two concerns we have $\Pi_l^0(\hat{h}) > \Pi_l^1(\hat{h})$, that is, transparency decreases the payoff of a judge from deviating from signal l.[18] Hence, it has to be effect (ii) that drives our results. It is indeed. We observe that both, for a judge with an only concern for bias and for a judge with an only concern for expertise, $\Pi_l^0(\hat{l}) > \Pi_l^1(\hat{l})$ when the judge's ability γ is not specially high.[19] Naturally, it implies that if the judge has a double concern, the same condition applies. That is, we find that when the judge is not specially able to interpret the law or the case is specially difficult, a judge that is thinking on passing a lenient sentence (after signal l) will prefer the state ω not to be revealed to the general public than it being publicly observed. In other words, a judge's payoff from following her signal is higher without transparency than with transparency.

However, note that this effect is also at work when we consider judges with an only concern and, as shown in Section 3, transparency had never a perverse effect there. Then, what is crucial for transparency to have a perverse effect only when we consider the two concerns together? The difference is that when the judge had an only concern for expertise, introducing transparency provoked a sharp decrease in her payoff from deviating and passing $\hat{h}$. As a result, a judge with an only concern for expertise stuck more to her signals in cases receiving media coverage than in those without it. In contrast to this, when we now consider a judge who also cares about bias and so, has an important incentive to avoid the lenient sentence, it occurs that transparency reduces more the judge's payoff from following her signal (and sending $\hat{l}$) than that from deviating (and passing $\hat{h}$). As a consequence, we obtain that media coverage of judicial cases can induce a judge to pass harsher sentences in cases receiving media attention than in those with no resonance beyond the courtroom.

4.2 The more general case

Let us now consider the more general case where the probability that the general public learns the true state after the judge's sentence takes any value in the interval $(0, 1)$. This could be the case, for example, when at the time to pass the sentence the judge does not know whether the sentence is going to attract media attention or not. Alternatively, it could illustrate a situation in which the general public can learn about the state ω from other sources (for example the citizens' own reading of a sentence) that the judge cannot control. For these situations, the next result shows that an increase in media coverage increases the severity of punishments. To show it, we first define $\sigma_l^\mu(\hat{l})^{Sup*}$ to be the supreme of all possible equilibrium strategy after signal l, $\sigma_l^\mu(\hat{l})^*$. That is, $\sigma_l^\mu(\hat{l})^{Sup*}$ denotes the judge's equilibrium strategy in which she follows signal l with a higher probability.

Proposition 5. *Let $\alpha_B^* < \alpha_B < \alpha_B^{**}$ and $\frac{1}{2} < \gamma < \tilde{\gamma}$. For $\mu \in (0, 1)$ we have that $\sigma_l^\mu(\hat{l})^{Sup*}$ decreases as μ increases.*

In line with the result in Proposition 4, we obtain that when the general public perception about the proportion of lenient judges in the system is neither too high nor too low, and the judge is not specially able to interpret the law or, alternatively, the case is specially difficult; a careerist judge with two concerns will pass harsher sentences the higher the probability that the general public learns about the correct sentence. In this sense, our conclusion that media coverage of judicial cases distorts judges' behavior and induces harsher punishments holds for any level of transparency.

[18] Note that the case with two concerns is simply a linear convex combination of the previous two cases.

[19] In particular, we require $\gamma < \bar{\gamma}$, with $\bar{\gamma} = \frac{\alpha_B + \alpha_W}{2\alpha_B + \alpha_W}$.

5 Conclusion

This paper analyzes the effects of media coverage of judicial cases -that increases the probability that the general public learns the correct sentence- on the incentives of a careerist judge to act on her information. The novelty of our approach is to consider a judge who seeks to signal both, expertise and absence of bias.

Our analysis of the judge's behavior in this case shows that the interplay between these two objectives produces unexpected results. The more relevant one is that media coverage can have perverse effects and induce the judge to pass more severe sentences. This result is in line with the empirical findings in Lim et al. (2015), who showed that newspaper coverage significantly increase sentence length, and with the casual observation discussed in the introduction of media publicized cases receiving specially harsh sentences.

An assumption in our analysis is that the two states of the world are equally likely. This is a simplifying assumption that however adds value to our results, as it guarantees that herding effects are not behind our finding that judges pass too much harsh sentences.[20] Allowing one state to be more likely than the other would just introduce an incentive for the judge to go for the popular belief, which would reinforce or counterbalance the judge's incentive to pass the harsh sentence, depending on whether the popular belief supports the harsh or the lenient punishment, respectively.[21] The other important assumption in our model is to consider a biased judge that always passes lenient sentences. A more sophisticated model would treat this player as a strategic agent, as in Morris (2001). He considers a two period agency game with a biased expert that wants the decision maker to take the expert's preferred action (suppose action $\hat{l}$). Though we agree that introducing this kind of considerations in our model would enrich our analysis, we predict that for the appropriate discount factor, we could always guarantee the result that the strategic biased type takes action $\hat{l}$, which is what we require.

The results in this paper have important policy implications. First, that the perception of a bias in the judicial system is not without consequences. In fact, our results suggest that when judges care about perceived biases, they may react trying to avoid the stigmatized action, which results in too much politically correct sentences. This argument helps explain why, for example, the proportion of jail inmates in US has sharply increased in the last thirty years.[22] Giving a twist to the model , it also helps explain why in the US black felons receive (nearly 20%) longer sentences than white men for similar crimes.[23,24]

[20]See Banerjee (1992), Scharfstein and Stein (1990), Heidhues and Lagerlöf (2003), Cummins and Nyman (2005) or Ottaviani and Sorensen (2006) for models of herding.

[21]See Andina-Díaz and García-Martínez (2015) for a formal analysis of a similar trade-off in a different setting.

[22]In 1970 the proportion of US prisoners was below one in 400; in 2010 it is one in 100. See "Too many laws, too many prisoners", *The Economists*, July 22nd, 2010.

[23]See "Racial gap in men's sentencing", *The Wall Street Journal*, February 14th, 2014.

[24]The twist to the model just requires a group of Americans to believe black defendants to be more likely to be guilty and the judicial system to be lenient with them, and judges to be concerned about this public perception. According to a study by Nazgol Ghandnoosh, it seems to be the case. He finds that "*When asked for numerical estimates of crime rates, whites attribute an exaggerated amount to people of color. And when asked to what degree various racial groups are "prone to violence", whites rank people of color as more violence-prone than their own race.*" He also refers to a study by Ted Chiricos and colleagues: "*They found that whites -though not blacks and Hispanics- who attributed higher proportions of violent crime, burglary, or robbery to blacks were significantly more likely to support these punitive policies.*" See "Race and Punishment: Racial Perceptions of Crime and Support for Punitive Policies", *The Sentencing Project*, September 3rd, 2014.

In this sense, law makers and legislators should be aware of this effect and introduce mechanisms to correct it. Second, that media coverage of sentencing practices may have pernicious effect and more precisely, it may induce judges to pass harsher sentences. In the paper we have shown that for this effect to occur, it has to be that a countries' justice is perceived as slightly biased and that the judge in charge of the case is not specially able or, alternatively, the case is particularly difficult. In these cases, law makers and higher instances in the judicial system should again be aware of the pernicious effect of mass media and try to counterbalance it, for example, assigning these cases to highly experience judges.

Beyond the courtroom, we consider that our model applies to any situation in which a career concern expert seeks to signal both expertise and bias. For example, a physician that seeks a promotion and does not want to be perceived as the type that prescribes too much artificial drugs; or a legislator that plans to run for reelection and wants to avoid being perceived as incompetent, etc. To any of these situations, our analysis shows that we have to be cautious with transparency, as learning the consequences of the expert's actions is not always good.

A Appendix

Appendix A contains the proofs and Appendix B analyzes the game with a strategic wise judge.

The sequence of the results proved in Appendix A is the following. We first prove Proposition 1 and then the rest of results, in the order of appearance in the text. The reason why we first prove Proposition 1 is that it characterizes the judge's behavior after signal h both, when the judge has two concerns and when she has an only concern (either for expertise or bias). Thus, the result in Proposition 1 simplifies the rest of the analyses in the text.

Prior to the proofs, we first formally define what an equilibrium strategy is in our game:

Definition 1. *Given $\mu \in [0,1]$, an equilibrium strategy $(\sigma_l^\mu(\hat{l})^*, \sigma_h^\mu(\hat{h})^*)$ is of the form:*

1. *If $\Delta_l^\mu[\overline{\sigma}_l^\mu(\hat{l}), \overline{\sigma}_h^\mu(\hat{h})] = \Delta_h^\mu[\overline{\sigma}_l^\mu(\hat{l}), \overline{\sigma}_h^\mu(\hat{h})] = 0$, then $(\overline{\sigma}_l^\mu(\hat{l}), \overline{\sigma}_h^\mu(\hat{h}))$ is an equilibrium strategy.*
2. *If $\Delta_l^\mu[\sigma_l^\mu(\hat{l}), \overline{\sigma}_h^\mu(\hat{h})] > 0$ (< 0) for all $\sigma_l^\mu(\hat{l}) \in [0,1]$, then $(\sigma_l^\mu(\hat{l}), \overline{\sigma}_h^\mu(\hat{h}))$ is an equilibrium strategy only if $\sigma_l^\mu(\hat{h}) = 1$ (0).*
3. *If $\Delta_h^\mu[\overline{\sigma}_l^\mu(\hat{l}), \sigma_h^\mu(\hat{h})] > 0$ (< 0) for all $\sigma_h^\mu(\hat{h}) \in [0,1]$, then $(\overline{\sigma}_l^\mu(\hat{l}), \sigma_h^\mu(\hat{h}))$ is an equilibrium strategy only if $\sigma_h^\mu(\hat{h}) = 1$ (0).*

where, given equations (1) − (7), we have:

$$\begin{aligned}\Delta_l^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] &= (1-\mu)\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] + \mu\Delta_l^1[\sigma_l(\hat{l}), \sigma_h(\hat{h})] \\ &= (1-\mu)\left(f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0], \lambda_{\bar{B}}[\hat{l},0]\right)\right) + \\ &\mu\left(\left(\gamma f(0,1) + (1-\gamma) f\left(\lambda_W[\hat{h},H],1\right)\right) - \left(\gamma f\left(\lambda_W[\hat{l},L], \lambda_{\bar{B}}[\hat{l},L]\right) + (1-\gamma) f\left(0, \lambda_{\bar{B}}[\hat{l},H]\right)\right)\right)\end{aligned} \tag{9}$$

$$\begin{aligned}\Delta_h^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] &= (1-\mu)\Delta_h^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] + \mu\Delta_h^1[\sigma_l(\hat{l}), \sigma_h(\hat{h})] \\ &= (1-\mu)\left(f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0], \lambda_{\bar{B}}[\hat{l},0]\right)\right) + \\ &\mu\left(\left(\gamma f\left(\lambda_W[\hat{h},H],1\right) + (1-\gamma) f(0,1)\right) - \left(\gamma f\left(0, \lambda_{\bar{B}}[\hat{l},H]\right) + (1-\gamma) f\left(\lambda_W[\hat{l},L], \lambda_{\bar{B}}[\hat{l},L]\right)\right)\right)\end{aligned} \tag{10}$$

Proof of Proposition 1

To prove the proposition, we need Claims (1)-(4).

Claim 1. $\lambda_W[\hat{h}, H] > \lambda_W[\hat{l}, L] \Longleftrightarrow (\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \alpha_B$.

Proof

We take $\lambda_W[\hat{h}, H]$ and $\lambda_W[\hat{l}, L]$ from Table 1. Now:
$\lambda_W[\hat{h}, H] > \lambda_W[\hat{l}, L] \Longleftrightarrow \frac{\lambda_W[\hat{h},H]}{\lambda_W[\hat{l},L]} > 1 \Longleftrightarrow \frac{\alpha_B+\alpha_W+\big((1-\gamma)\sigma_h(\hat{l})+\gamma(1-\sigma_l(\hat{h}))\big)\alpha_N}{\alpha_W+\big(\gamma(1-\sigma_h(\hat{l}))+(1-\gamma)\sigma_l(\hat{h})\big)\alpha_N} > 1$
$\Longleftrightarrow \alpha_B - \sigma_l(\hat{h})\alpha_N + \sigma_h(\hat{l})\alpha_N > 0 \Longleftrightarrow (\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \alpha_B$. ♦

Claim 2. $\lambda_W[\hat{h}, 0] > \lambda_W[\hat{l}, 0] \Longleftrightarrow (\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \alpha_B$.

Proof

We take $\lambda_W[\hat{h}, 0]$ and $\lambda_W[\hat{l}, 0]$ from Table 1. Now:
$\lambda_W[\hat{h}, 0] > \lambda_W[\hat{l}, 0] \Longleftrightarrow \frac{\lambda_W[\hat{h},0]}{\lambda_W[\hat{l},0]} > 1 \Longleftrightarrow \frac{2\alpha_B+\alpha_W+(1-\sigma_l(\hat{h})+\sigma_h(\hat{l}))\alpha_N}{\alpha_W+(1+\sigma_l(\hat{h})-\sigma_h(\hat{l}))\alpha_N} > 1$
$\Longleftrightarrow \alpha_B - \sigma_l(\hat{h})\alpha_N + \sigma_h(\hat{l})\alpha_N > 0 \Longleftrightarrow (\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \alpha_B$. ♦

Claim 3. $\lambda_{\bar{B}}[\hat{l}, H] < \lambda_{\bar{B}}[\hat{l}, L]$ *in equilibrium.*

Proof

$\lambda_{\bar{B}}[\hat{l}, H] < \lambda_{\bar{B}}[\hat{l}, L] \Longleftrightarrow 1 - \lambda_B[\hat{l}, H] < 1 - \lambda_B[\hat{l}, L] \Longleftrightarrow \lambda_B[\hat{l}, H] > \lambda_B[\hat{l}, L]$, where $\lambda_B[\hat{l}, H]$ and $\lambda_B[\hat{l}, L]$ are given by Table 2. Now:
$\lambda_B[\hat{l}, H] > \lambda_B[\hat{l}, L] \Longleftrightarrow \frac{\lambda_B[\hat{l},H]}{\lambda_B[\hat{l},L]} > 1 \Longleftrightarrow \frac{\alpha_B+\alpha_W+\big((1-\gamma)\sigma_h(\hat{l})+\gamma(1-\sigma_l(\hat{h}))\big)\alpha_N}{\alpha_B+\big(\gamma\sigma_h(\hat{l})+(1-\gamma)(1-\sigma_l(\hat{h}))\big)\alpha_N} > 1$
$\Longleftrightarrow \alpha_W - \alpha_N + \sigma_l(\hat{h})\alpha_N + \sigma_h(\hat{l})\alpha_N + 2\gamma\alpha_N - 2\sigma_l(\hat{h})\gamma\alpha_N - 2\sigma_h(\hat{l})\gamma\alpha_N > 0$
$\Longleftrightarrow \alpha_W + (2\gamma - 1)((1 - \sigma_l(\hat{h})) - \sigma_h(\hat{l})))\alpha_N > 0$.

On the one hand $2\gamma - 1 > 0$, since $\gamma > \frac{1}{2}$. On the other hand, in a non-perverse equilibrium, $\sigma_h(\hat{l}) < (1 - \sigma_l(\hat{h}))$. Then, the expression above is positive. ♦

Claim 4. $\Delta_l^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] < \Delta_h^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})]$.

Proof

Since $\Delta_s^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = (1-\mu)\Delta_s^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] + \mu\Delta_s^1[\sigma_l(\hat{l}), \sigma_h(\hat{h})]$ for any $s \in \{l, h\}$, and $\Delta_l^0 = \Delta_h^0$, we just have to focus on equations Δ_l^1 and Δ_h^1, which are the second adding in equations (9) and (10), respectively.

Now,
$\Delta_l^1[\sigma_l(\hat{l}), \sigma_h^\mu(\hat{h})] < \Delta_h^1[\sigma_l(\hat{l}), \sigma_h(\hat{h})] \Longleftrightarrow$
$(\gamma - (1-\gamma))\left(f(0,1) - f\left(\lambda_W[\hat{h}, H], 1\right)\right) < (\gamma - (1-\gamma))\left(f\left(\lambda_W[\hat{l}, L], \lambda_{\bar{B}}[\hat{l}, L]\right) - f\left(0, \lambda_{\bar{B}}[\hat{l}, H]\right)\right)$
$\Longleftrightarrow f\left(\lambda_W[\hat{h}, H], 1\right) + f\left(\lambda_W[\hat{l}, L], \lambda_{\bar{B}}[\hat{l}, L]\right) > f(0,1) + f\left(0, \lambda_{\bar{B}}[\hat{l}, H]\right)$.

Since $f\left(\lambda_W[\hat{h}, H], 1\right) > f(0,1)$ and by Claim 3 we know that $f\left(\lambda_W[\hat{l}, L], \lambda_{\bar{B}}[\hat{l}, L]\right) > f\left(0, \lambda_{\bar{B}}[\hat{l}, H]\right)$, the inequality above always holds. ♦

Now, we can prove Proposition 1.

From Claim 4, we know that in equilibrium it cannot be that the judge lies with positive probability in the two information sets. We prove it by contradiction. To this, suppose $\sigma_l(\hat{h})^* > 0$ and $\sigma_h(\hat{l})^* > 0$.

Use Definition 1. Now, since $\sigma_h(\hat{h})^* < 1$, then $\Delta_h^\mu \leq 0$. Additionally, since $\sigma_l(\hat{l})^* < 1$, then $\Delta_l^\mu \geq 0$, which contradicts Claim 4. Then, the equilibrium strategy profile is either $(\sigma_l(\hat{h})^* \geq 0, \sigma_h(\hat{l})^* = 0)$ or $(\sigma_l(\hat{h})^* = 0, \sigma_h(\hat{l})^* \geq 0)$.

Next we show that $\sigma_h(\hat{l})^* > 0$ is not possible. Again, we prove it by contradiction. Suppose $\sigma_h(\hat{l})^* > 0$, then $\sigma_l(\hat{h})^* = 0$. Consequently, by Claims 1 and 2, $\lambda_W[\hat{h}, H] > \lambda_W[\hat{l}, L]$ and $\lambda_W[\hat{h}, 0] > \lambda_W[\hat{l}, 0]$. Substituting in equation (10) we obtain $\Delta_h^\mu > 0$. But $\Delta_h^\mu > 0$ implies $\sigma_h(\hat{h})^* = 1$, which contradicts $\sigma_h(\hat{l})^* > 0$. As a result, in equilibrium $\sigma_h(\hat{l})^* = 0$, i.e. $\sigma_h(\hat{h})^* = 1$. ■

Proof of Lemma 1

Here $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X)) = \lambda_W(a,X)$ and $\mu = 0$. In this case, expressions 9 and 10 simplify to:

$$\Delta_l^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \Delta_h^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \lambda_W[\hat{h}, 0] - \lambda_W[\hat{l}, 0].$$

After some algebra, we obtain:

$\Delta_l^0 = \Delta_h^0 = \lambda_W[\hat{h}, 0] - \lambda_W[\hat{l}, 0] = \frac{\alpha_W}{\alpha_W + (\sigma_h(\hat{h}) + \sigma_l(\hat{h}))\alpha_N} - \frac{\alpha_W}{2\alpha_B + \alpha_W + (\sigma_h(\hat{l}) + \sigma_l(\hat{l}))\alpha_N} \geq 0 \iff \frac{\alpha_B}{\alpha_N} + \sigma_h(\hat{l}) \geq \sigma_l(\hat{h})$.

There are two cases:

(i) If $\alpha_B \geq \alpha_N$, then $\frac{\alpha_B}{\alpha_N} + \sigma_h(\hat{l}) > 1$ and consequently $\Delta_l^0 = \Delta_h^0 > 0$ for all $\sigma_h(\hat{l})$ and $\sigma_l(\hat{h})$. Hence, in the unique equilibrium, $\sigma_l(\hat{l})^* = 0$ and $\sigma_h(\hat{h})^* = 1$.

(ii) If $\alpha_B < \alpha_N$, then $\frac{\alpha_B}{\alpha_N} < 1$ and consequently $\Delta_l^0 = \Delta_h^0 = 0$ if $\frac{\alpha_B}{\alpha_N} + \sigma_h(\hat{l}) = \sigma_l(\hat{h})$. Therefore, $\frac{\alpha_B}{\alpha_N} + \sigma_h(\hat{l})^* = \sigma_l(\hat{h})^*$. Note that in this case, the profile $\sigma_h(\hat{l}) = 0$ and $\sigma_l(\hat{h}) = 1$ cannot be an equilibrium because this implies that $\Delta_l^0 = \Delta_h^0 < 0$. The profile $\sigma_h(\hat{l}) = 1$ and $\sigma_l(\hat{h}) = 0$ cannot be an equilibrium because this implies that $\Delta_l^0 = \Delta_h^0 > 0$. Additionally, by Remark 1 in equilibrium $\sigma_h^*(\hat{h}) = 1$ and consequently $\sigma_l(\hat{l})^* = 1 - \frac{\alpha_B}{\alpha_N}$.

Proof of Lemma 2

First, note that

$$\Delta_l^1 = (1-\gamma)\lambda_W[\hat{h}, H] - \gamma\lambda_W[\hat{l}, L] = \frac{(1-\gamma)\alpha_W}{\alpha_W + (\gamma\sigma_h(\hat{h}) + (1-\gamma)\sigma_l(\hat{h}))\alpha_N} - \frac{\gamma\alpha_W}{\alpha_B + \alpha_W + ((1-\gamma)\sigma_h(\hat{l}) + \gamma\sigma_l(l))\alpha_N}$$

with $\frac{\partial \Delta_l^1}{\partial \sigma_l(\hat{l})} > 0$. Consequently, there is either one root for $\sigma_l(\hat{l})$ or none. In other words, the equilibrium is unique. In addition, by Remark 1, $\sigma_h(\hat{h})^* = 1$.

Next, we evaluate Δ_l^1 at $\sigma_l(\hat{l}) = 0$ and $\sigma_l(\hat{l}) = 1$, with $\sigma_h(\hat{h}) = 1$. After some algebra we obtain:

$\Delta_l^1[\sigma_l(\hat{l}) = 1, \sigma_h(\hat{h}) = 1] \leq 0 \iff \alpha_B \leq \bar{\alpha}_B^0 = \frac{(2\gamma-1)(\gamma+\alpha_W-\gamma\alpha_W)}{\gamma^2+(\gamma-1)^2}$. Thus, the function $\Delta_l^1[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1]$ is negative for all $\sigma_l(\hat{l})$, consequently $\sigma_l(\hat{l})^* = 1$.

$\Delta_l^1[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] \geq 0 \iff \alpha_B \geq \bar{\alpha}_B^1 = \gamma - \alpha_W(1-\gamma)$. The function $\Delta_l^1[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1]$ is positive for all $\sigma_l(\hat{l})$, consequently $\sigma_l(\hat{l})^* = 0$.

When $\bar{\alpha}_B^0 < \alpha_B < \bar{\alpha}_B^1$, note that $\Delta_l^1[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] < 0$ and $\Delta_l^1[\sigma_l(\hat{l}) = 1, \sigma_h(\hat{h}) = 1] > 0$, which implies that there is one root,

$\Delta_l^1[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1] = 0 \iff \sigma_l(\hat{l}) = \frac{\gamma(\alpha_W+\alpha_N)+(\gamma-1)(\alpha_B+\alpha_W)}{2\gamma\alpha_N(1-\gamma)}$. ■

Proof of Lemma 3

Note that $\Delta_l^\mu = (1-\mu)\,\Delta_l^0 + \mu\Delta_l^1$, with $\frac{\partial \Delta_l^0}{\partial \sigma_l(\hat{l})} > 0$ and $\frac{\partial \Delta_l^1}{\partial \sigma_l(\hat{l})} > 0$. Then, $\frac{\partial \Delta_l^\mu}{\partial \sigma_l(\hat{l})} > 0$. That is to say, the equilibrium is unique.

Now, let us denote by $\bar{\sigma}_l^0(\hat{l})$ and $\bar{\sigma}_l^1(\hat{l})$ the interior solutions to equations $\Delta_l^0 = 0$ and $\Delta_l^1 = 0$, respectively. From Lemmas 1 and 2, we know they are $\bar{\sigma}_l^0(\hat{l}) = 1 - \frac{\alpha_B}{\alpha_N}$ and $\bar{\sigma}_l^1(\hat{l}) = \frac{\gamma(\alpha_W+\alpha_N)+(\gamma-1)(\alpha_B+\alpha_W)}{2\gamma\alpha_N(1-\gamma)}$.

Next, note that $\bar{\sigma}_l^0(\hat{l}) < \bar{\sigma}_l^1(\hat{l})$

$\iff 1 - \frac{\alpha_B}{\alpha_N} - \frac{\gamma(\alpha_W+\alpha_N)+(\gamma-1)(\alpha_B+\alpha_W)}{2\gamma\alpha_N(1-\gamma)} < 0$

$\iff -\frac{(2\gamma-1)(\alpha_W(1-\gamma)\alpha_B+\gamma\alpha_N)}{2\gamma\alpha_N(1-\gamma)} < 0$, which is always the case.

Finally, note that $\Delta_l^0 < 0$ if $\sigma_l(\hat{l}) < \bar{\sigma}_l^0(\hat{l})$ and $\Delta_l^0 > 0$ if $\sigma_l(\hat{l}) > \bar{\sigma}_l^0(\hat{l})$. Similarly, $\Delta_l^1 < 0$ if $\sigma_l(\hat{l}) < \bar{\sigma}_l^1(\hat{l})$ and $\Delta_l^1 > 0$ if $\sigma_l(\hat{l}) > \bar{\sigma}_l^1(\hat{l})$. Then, if $0 < \sigma_l^\mu(\hat{l})^* < 1$, it must be that $\sigma_l^\mu(\hat{l})^* \in \left[\bar{\sigma}_l^0(\hat{l}), \bar{\sigma}_l^1(\hat{l})\right]$. Last, since $\Delta_l^\mu = (1-\mu)\,\Delta_l^0 + \mu\Delta_l^1$, with $\Delta_l^0 > 0$ and $\Delta_l^1 < 0$ in the interval $\left[\bar{\sigma}_l^0(\hat{l}), \bar{\sigma}_l^1(\hat{l})\right]$, then $\sigma_l^\mu(\hat{l})^*$ has to be increasing in μ. ■

Proof of Proposition 2

Note that $\Delta_l^\mu = (1-\mu)\Delta_l^0 + \mu\Delta_l^1$. Then, it is sufficient to prove that there exists an $\overset{\sup}{\alpha}_B$ such that for any $\alpha_B > \overset{\sup}{\alpha}_B$, both $\Delta_l^0 > 0$ and $\Delta_l^1 > 0$. The following two claims show these results.

Claim 5. *For any strictly increasing function $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X))$, there always exists an $\breve{\alpha}_B < 1$ such that for any $\alpha_B > \breve{\alpha}_B$, $\Delta_l^0 > 0$.*

Proof

Note that $\Delta_l^0 = f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0], \lambda_{\bar{B}}[\hat{l},0]\right)$.

A sufficient condition for $\Delta_l^0 > 0$ is $\lambda_W[\hat{h},0] > \lambda_W[\hat{l},0] \iff (\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \alpha_B$ (from Claim 2). Now, since $(\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < 1$, there always exists an $\breve{\alpha}_B < 1$ for which $(\sigma_l(\hat{h}) - \sigma_h(\hat{l}))\alpha_N < \breve{\alpha}_B$. Consequently, for any $\alpha_B > \breve{\alpha}_B$, $\Delta_l^0 > 0$. ♦

Claim 6. *For any strictly increasing function $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X))$, there always exists an $\breve{\alpha}_B < 1$ such that for any $\alpha_B > \breve{\alpha}_B$, $\Delta_l^1 > 0$.*

Proof

Note that $\Delta_l^1 = \Pi_l^1(\hat{h}) - \Pi_l^1(\hat{l})$, with $\Pi_l^1(\hat{h}) = \gamma f\,(0,1) + (1-\gamma) f\left(\lambda_W[\hat{h},H],1\right)$ and $\Pi_l^1(\hat{l}) = \gamma f\left(\lambda_W[\hat{l},L], \lambda_{\bar{B}}[\hat{l},L]\right) + (1-\gamma) f\left(0, \lambda_{\bar{B}}[\hat{l},H]\right)$.

With, $\Pi_l^1(\hat{h}) = \gamma f\,(0,1) + (1-\gamma) f\left(\lambda_W[\hat{h},H],1\right) > \gamma f\,(0,1) + (1-\gamma) f\,(0,1) = f\,(0,1) > 0$.

On the other hand, the function $\Pi_l^1(\hat{l}) = \gamma f\left(\lambda_W[\hat{l},L], \lambda_{\bar{B}}[\hat{l},L]\right) + (1-\gamma) f\left(0, \lambda_{\bar{B}}[\hat{l},H]\right)$ is decreasing in α_B (with α_W constant and also with α_B constant). In addition $lim_{\alpha_B \to 1}\Pi_l^1(\hat{l}) \to 0$, because all the beliefs go to zero as α_B goes to one.

Thus, there always exists an $\breve{\alpha}_B < 1$ such that for any $\alpha_B > \breve{\alpha}_B$, $\Pi_l^1(\hat{h}) > \Pi_l^1(\hat{l})$ and consequently, $\Delta_l^1 > 0$. ♦

Finally, let us denote $\overset{\sup}{\alpha}_B = \max\{\breve{\alpha}_B, \breve{\alpha}_B\}$. Now, since $\Delta_l^\mu = (1-\mu)\Delta_l^0 + \mu\Delta_l^1 > 0$, and from Claims 5 and 6 we know that for any $\alpha_B > \overset{\sup}{\alpha}$, $\Delta_l^0 > 0$ and $\Delta_l^1 > 0$, the proof follows. ■

Proof of Proposition 3

We first prove two claims.

Claim 7. *For any strictly increasing function $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X))$ and for any $\varepsilon_1>0$, there always exits an $\alpha_B^{(\varepsilon_1)}>0$ such that for any $\alpha_B<\alpha_B^{(\varepsilon_1)}$, $|\Delta_l^0|<\varepsilon_1$.*

Proof

Note that $\Delta_l^0 = f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0],\lambda_{\bar{B}}[\hat{l},0]\right)$.

On the one hand, it is straightforward to show that $\frac{\lambda_{\bar{B}}[\hat{l},0]}{\alpha_B}<0$, with $lim_{\alpha_B\to 0}\lambda_{\bar{B}}[\hat{l},0]\to 1$.

On the other hand, $\lambda_W[\hat{h},0]\leq\lambda_W[\hat{l},0]\iff(\sigma_l(\hat{h})-\sigma_h(\hat{l}))\alpha_N\geq\alpha_B$ (from Claim 2). Now, using Remark 1, i.e. $\sigma_h(\hat{l})=0$, the inequality above holds for α_B sufficiently small. Additionally, even if $\sigma_l(\hat{h})=0$, it is straightforward to show that $\lambda_W[\hat{l},0]\to\lambda_W[\hat{h},0]$ when $\alpha_B\to 0$.

Consequently, $\Delta_l^0\to 0$ when $\alpha_B\to 0$. Thus, for any $\varepsilon_1>0$, there exits an $\alpha_B^{(\varepsilon_1)}>0$ such that for any $\alpha_B<\alpha_B^{(\varepsilon_1)}$, $|\Delta_l^0|<\varepsilon_1$. ♦

Claim 8. *For any strictly increasing function $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X))$, there exits $\alpha_B^{'}>0$ and $\overset{\sup}{\gamma}<1$ such that for any $\alpha_B<\alpha_B^{'}$, and $\gamma>\overset{\sup}{\gamma}$, $\Delta_l^1<0$.*

Proof

Note that $\Delta_l^1=\gamma f(0,1)+(1-\gamma)f\left(\lambda_W[\hat{h},H],1\right)-\left(\gamma f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right)+(1-\gamma)f\left(0,\lambda_{\bar{B}}[\hat{l},H]\right)\right)$.

First, we show that $f(0,1)<f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right)$ if α_B is sufficiently small. On the one hand, we can easily observe that $lim_{\alpha_B\to 0}\lambda_{\bar{B}}[\hat{l},L]\to 1$. On the other hand, $\lambda_W[\hat{l},L]=\frac{\alpha_W}{\alpha_B+\alpha_W+((1-\gamma)\sigma_h(\hat{l})+\gamma\sigma_l(l))\alpha_N}>\alpha_W>0$. In addition, it is straightforward to show that $\lambda_W[\hat{l},L]$ increases when α_B decreases. Thus, there always exits $\alpha_B^{'}>0$ such that, if $\alpha_B<\alpha_B^{'}$, then $f(0,1)<f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right)$. In that case, if γ is high enough, then $\Delta_l^1<0$. Consequently, for any $\alpha_B^{'}$ there is $\overset{\sup}{\gamma}<1$, such that for all $\gamma<\overset{\sup}{\gamma}$ and $\alpha_B<\alpha_B^{'}$, $\Delta_l^1<0$. ♦

Finally, there always exists $\alpha_B^{'}>0$ and $\gamma^{'}<1$ such that for any $\alpha_B<\alpha_B^{'}$ and $\gamma<\gamma^{'}$, $\Delta_l^1<0$. From Claim 7, if we take $\alpha_B^{(\varepsilon_1)}$ such that $\varepsilon_1<|\Delta_l^1|$, and let $\overset{\inf}{\alpha}_B=\min\{\alpha_B^{(\varepsilon_1)},\alpha_B^{'}\}$, then if $\alpha_B<\overset{\inf}{\alpha}_B$ and $\gamma>\overset{\sup}{\gamma}$, $\Delta_l^\mu=(1-\mu)\Delta_l^0+\mu\Delta_l^1<0$ for any μ. Consequently, $\sigma_l^{\bar{\mu}}(\hat{l})^*=1$ for all μ. ■

Lemma 4. *Consider $\mu=0$. For any increasing function $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X))$, in equilibrium $\sigma_l^0(\hat{l})^*<1$.*

Proof. We prove it by contradiction. Suppose $\sigma_l^0(\hat{l})^*=1$. Then $\Delta_l^0<0$. Now, let us evaluate Δ_l^0 at $\sigma_l^0(\hat{l})^*=1$ and $\sigma_h^0(\hat{h})^*=1$ (using Remark 1). We obtain: $\Delta_l^0[\sigma_l(\hat{l})=1,\sigma_h(\hat{h})=1]=f\left(\frac{\alpha_W}{\alpha_W+\alpha_N},1\right)-f\left(\frac{\alpha_W}{2\alpha_B+\alpha_W+\alpha_N},1-\frac{2\alpha_B}{2\alpha_B+\alpha_W+\alpha_N}\right)>0$. A contradiction. ■ □

Proof of Proposition 4

The following Lemma 5 proves Proposition 4 for $\mu=0$, and Lemma 6 proves it for $\mu=1$.

Lemma 5. *Consider $\mu=0$. In equilibrium:*[25]

1) If $\alpha_B<\alpha_B^{**}$, then $\left(\sigma_l^0(\hat{l})^*=1-\frac{\alpha_B(1-\alpha_B+\alpha_W)}{(\alpha_W-\alpha_B)(1-\alpha_W-\alpha_B)},\sigma_h^0(\hat{h})^*=1\right)$.

[25] As $\Delta_l^0[\sigma_l(\hat{l}),\sigma_h(\hat{h})]=\Delta_h^0[\sigma_l(\hat{l}),\sigma_h(\hat{h})]$, there are multiple equilibria. From the proof of Proposition 4, it is straightforward to derive that if $\alpha_B<\alpha_B^{**}$, then $\sigma_l^0(\hat{h})^*=\frac{\alpha_B(1-\alpha_B+\alpha_W)}{(\alpha_W-\alpha_B)(1-\alpha_W-\alpha_B)}+\sigma_h^0(\hat{l})^*$. However, by Remark 1, we consider $\sigma_h^0(\hat{l})^*=0$.

2) If $\alpha_B \geq \alpha_B^{**}$, then $\left(\sigma_l^0(\hat{l})^* = 0, \sigma_h^0(\hat{h})^* = 1\right)$, with $\alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$.

Proof.

First, we state the following four claims.

Claim 9. $\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \Delta_h^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = 0 \iff \sigma_l(\hat{l}) = \sigma_h(\hat{h}) - \frac{2\alpha_B\alpha_W + \alpha_B\alpha_N}{(\alpha_W - \alpha_B)\alpha_N}$.

Proof.

From equations (9) and (10),

$$\begin{aligned}\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \Delta_h^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] &= f\left(\lambda_W[\hat{h}, 0], 1\right) - f\left(\lambda_W[\hat{l}, 0], \lambda_{\bar{B}}[\hat{l}, 0]\right) \\ &= \left(\lambda_W[\hat{h}, 0] + 1\right) - \left(\lambda_W[\hat{l}, 0] + \lambda_{\bar{B}}[\hat{l}, 0]\right) \\ &= \frac{\alpha_W}{\alpha_W + (\sigma_h(\hat{h}) + \sigma_l(\hat{h}))\alpha_N} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W + (\sigma_h(\hat{l}) + \sigma_l(\hat{l}))\alpha_N}\end{aligned}$$

Where, after some algebra

$\frac{\alpha_W}{\alpha_W + (\sigma_h(\hat{h}) + \sigma_l(\hat{h}))\alpha_N} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W + (\sigma_h(\hat{l}) + \sigma_l(\hat{l}))\alpha_N} = 0$

$\iff \sigma_l(\hat{l}) = \sigma_h(\hat{h}) - \frac{2\alpha_B\alpha_W + \alpha_B\alpha_N}{(\alpha_W - \alpha_B)\alpha_N}$. ♦

Claim 10. *If $\alpha_B > \frac{1}{2}\alpha_W$, then $\sigma_l^0(\hat{l})^* = 0$ and $\sigma_h^0(\hat{h})^* = 1$.*

Proof.

As $\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \Delta_h^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})] = \frac{\alpha_W}{\alpha_W + (\sigma_h(\hat{h}) + \sigma_l(\hat{h}))\alpha_N} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W + (\sigma_h(\hat{l}) + \sigma_l(\hat{l}))\alpha_N}$, if $2\alpha_B > \alpha_W$, then $\Delta_l^0 = \Delta_h^0 > 0$ and consequently $\sigma_l^{\mu=0}(\hat{l})^* = 0$ and $\sigma_h^0(\hat{h})^* = 1$. ♦

Claim 11. $\frac{\partial \Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h})]}{\partial \sigma_l(\hat{l})} > 0$ *if $\alpha_B < \frac{1}{2}\alpha_W$.*

Proof.

$\frac{\partial \Delta_l^\mu}{\partial \sigma_l(\hat{l})} = \frac{\alpha_W\alpha_N}{\left(\alpha_W + \alpha_N + \alpha_N\sigma_h(\hat{h}) - \alpha_N\sigma_l(\hat{l})\right)^2} + \frac{(\alpha_W - 2\alpha_B)\alpha_N}{\left(2\alpha_B + \alpha_W + \alpha_N\sigma_h(\hat{l}) + \alpha_N\sigma_l(\hat{l})\right)^2} > 0$. ♦

As $\alpha_B < \frac{1}{2}\alpha_W$, the numerator of the second adding is also positive $(\alpha_W - 2\alpha_B) > 0$.

Claim 12. $\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] > 0 \iff \alpha_B > \alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$ *where* $\alpha_B^{**} < \frac{1}{2}\alpha_W$.

Proof.

$\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] = \frac{\alpha_W}{\alpha_W + 2\alpha_N} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W}$. Now, substituting $\alpha_N = (1 - \alpha_B - \alpha_W)$ we obtain $\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] = \frac{\alpha_W}{\alpha_W + 2(1 - \alpha_B - \alpha_W)} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W}$, where $\frac{\alpha_W}{\alpha_W + 2(1 - \alpha_B - \alpha_W)} + \frac{2\alpha_B - \alpha_W}{2\alpha_B + \alpha_W} > 0$

$\iff \frac{-2\alpha_B^2 + \alpha_B\alpha_W + 2\alpha_B + \alpha_W^2 - \alpha_W}{(2 - 2\alpha_B - \alpha_W)(2\alpha_B + \alpha_W)} > 0$

$\iff -2\alpha_B^2 + \alpha_B\alpha_W + 2\alpha_B + \alpha_W^2 - \alpha_W > 0$.

Now, the roots of $-2\alpha_B^2 + \alpha_B\alpha_W + 2\alpha_B + \alpha_W^2 - \alpha_W = 0$ are $\frac{1}{4}\left(2 + \alpha_W \pm \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$. It is straightforward to show that $\frac{1}{4}\left(2 + \alpha_W + \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right) > 1$. Thus, $\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] > 0 \iff \alpha_B > \alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$.

Last, note that $\alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right) < \frac{1}{2}\alpha_W \iff 2 - \alpha_W < \sqrt{9\alpha_W^2 - 4\alpha_W + 4}$, and $\sqrt{9\alpha_W^2 - 4\alpha_W + 4} > \sqrt{\alpha_W^2 - 4\alpha_W + 4} = \sqrt{(2 - \alpha_W)^2} = 2 - \alpha_W$. ♦

Claim 13. $\Delta_l^0[\sigma_l(\hat{l}) = 1, \sigma_h(\hat{h}) = 1] > 0$

Proof.

It is proved in the proof of Lemma 4. ⧫

By Remark 1, $\sigma_h^0(\hat{h})^* = 1$.

First, we prove that if $\alpha_B > \alpha_B^{**}$, then $\sigma_l^0(\hat{l})^* = 0$. Note that if $\alpha_B \in \left(\alpha_B^{**}, \frac{1}{2}\alpha_W\right)$, then $\Delta_l^0[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1] > 0$ for all $\sigma_l(\hat{l})$ because $\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] > 0$ (by Claim 12) and Δ_l^0 is increasing in $\sigma_l(\hat{l})$ (by Claim 11). Consequently, $\sigma_l^0(\hat{l})^* = 0$ if $\alpha_B \in \left(\alpha_B^{**}, \frac{1}{2}\alpha_W\right)$. In addition, when $\alpha_B > \frac{1}{2}\alpha_W$, by Claim 10, $\sigma_l^0(\hat{l})^* = 0$.

Finally, if $\alpha_B < \alpha_B^{**}$, then $\Delta_l^0[\sigma_l(\hat{l}) = 0, \sigma_h(\hat{h}) = 1] < 0$ by Claim 12, and $\Delta_l^0[\sigma_l(\hat{l}) = 1, \sigma_h(\hat{h}) = 1] > 0$ by Claim 13. As function Δ_l^0 is increasing in $\sigma_l(\hat{l})$ (see Claim 11), there is one root, that is the equilibrium strategy. By Claim 9, it is $\sigma_l(\hat{l}) = \sigma_h(\hat{h}) - \frac{2\alpha_B\alpha_W + \alpha_B\alpha_N}{(\alpha_W - \alpha_B)\alpha_N}$. ⧫

Lemma 6. *With $\mu = 1$, if $\alpha_B^* < \alpha_B < \alpha_B^{**}$ and $\frac{1}{2} < \gamma < \tilde{\gamma}$, then there is a unique equilibrium in which $\left(\sigma_l^{\mu=1}(\hat{l})^* = 0, \sigma_h^{\mu=1}(\hat{h})^* = 1\right)$. Where, $\alpha_B^* = \frac{\alpha_W\left(2+\alpha_W+\alpha_W^2-2\sqrt{2}\sqrt{\alpha_W(\alpha_W+1)}\right)}{(\alpha_W+2)^2}$, $\alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$ and $\tilde{\gamma} = \frac{\left(\alpha_B^2-\alpha_W^2\right)}{2\sqrt{2\alpha_B\alpha_W}-(2\alpha_B+\alpha_W)(1-\alpha_B+\alpha_W)}$.*

Proof.

By Proposition 1, if $\mu = 1$, then $\sigma_h^1(\hat{h})^* = 1$.

From equation (9) and Tables 1 and 2,

$$\begin{aligned}
\Delta_l^1[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1] &= \left(\gamma f(0,1) + (1-\gamma) f\left(\lambda_W[\hat{h}, H], 1\right)\right) - \left(\gamma f\left(\lambda_W[\hat{l}, L], \lambda_{\bar{B}}[\hat{l}, L]\right) + (1-\gamma) f\left(0, \lambda_{\bar{B}}[\hat{l}, H]\right)\right) \\
&= \left(\gamma + (1-\gamma)\left(\lambda_W[\hat{h}, H] + 1\right)\right) - \left(\gamma\left(\lambda_W[\hat{l}, L] + \left(1 - \lambda_B[\hat{l}, L]\right)\right) + (1-\gamma)\left(1 - \lambda_B[\hat{l}, H]\right)\right) \\
&= (1-\gamma)\lambda_W[\hat{h}, H] - \gamma\left(\lambda_W[\hat{l}, L] - \lambda_B[\hat{l}, L]\right) + (1-\gamma)\lambda_B[\hat{l}, H] \\
&= \frac{(1-\gamma)\alpha_W}{\alpha_W + (\gamma + (1-\gamma)\sigma_l(\hat{h}))\alpha_N} - \frac{\gamma(\alpha_W - \alpha_B)}{\alpha_B + \alpha_W + (\gamma\sigma_l(\hat{l}))\alpha_N} + \frac{(1-\gamma)\alpha_B}{\alpha_B + ((1-\gamma)\sigma_l(\hat{l}))\alpha_N}
\end{aligned}$$

If the above expression is greater than zero, then in equilibrium $\sigma_l(\hat{l})^* = 0$ when $\mu = 1$. Clearly, the expression will be positive if $\alpha_B > \alpha_W$. In that case, the judge always takes action $\hat{h}$ and $\sigma_l(\hat{h})^* = 1$. Therefore, hereafter we consider that $\alpha_B < \alpha_W$.

Note that the denominators of the three fractions of Δ_l^1 are greater than zero, thus,

$$\begin{aligned}
&\frac{(1-\gamma)\alpha_W}{\alpha_W + \left(\gamma + (1-\gamma)\sigma_l(\hat{h})\right)\alpha_N} + \frac{\gamma(\alpha_B - \alpha_W)}{\alpha_B + \alpha_W + \left(\gamma(1-\sigma_l(\hat{h}))\right)\alpha_N} + \frac{(1-\gamma)\alpha_B}{\alpha_B + \left((1-\gamma)(1-\sigma_l(\hat{h}))\right)\alpha_N} > 0 \\
&\iff (1-\gamma)\alpha_W\left(\alpha_B + \alpha_W + \left(\gamma(1-\sigma_l(\hat{h}))\right)\alpha_N\right)\left(\alpha_B + \left((1-\gamma)(1-\sigma_l(\hat{h}))\right)\alpha_N\right) \\
&\quad + \gamma(\alpha_B - \alpha_W)\left(\alpha_W + \left(\gamma + (1-\gamma)\sigma_l(\hat{h})\right)\alpha_N\right)\left(\alpha_B + \left((1-\gamma)(1-\sigma_l(\hat{h}))\right)\alpha_N\right) \\
&\quad + (1-\gamma)\alpha_B\left(\alpha_W + \left(\gamma + (1-\gamma)\sigma_l(\hat{h})\right)\alpha_N\right)\left(\alpha_B + \alpha_W + \left(\gamma(1-\sigma_l(\hat{h}))\right)\alpha_N\right) > 0.
\end{aligned}$$

The expression above is a second-degree polynomial in $\sigma_l(\hat{h})$ and can be rewritten as:

$$\begin{aligned}
p(\sigma_l(\hat{h})) &= \sigma_l(\hat{h})^2\left(2\gamma\alpha_N^2(1-\gamma)^2(\alpha_W - \alpha_B)\right) \qquad (11) \\
&\quad + \sigma_l(\hat{h})\left(\alpha_N(1-\gamma)\left(-4\gamma^2\alpha_N\alpha_B + 4\gamma^2\alpha_N\alpha_W - 4\gamma\alpha_B\alpha_W + 2\gamma\alpha_N\alpha_B + 2\gamma\alpha_W^2 - 3\gamma\alpha_N\alpha_W + \alpha_B^2 - \alpha_W^2\right)\right) \\
&\quad + \gamma(\alpha_W + \gamma\alpha_N)(\alpha_B - \alpha_N(\gamma-1))(\alpha_B - \alpha_W) - \alpha_W(\gamma-1)(\alpha_B - \alpha_N(\gamma-1))(\alpha_B + \alpha_W + \gamma\alpha_N) \\
&\qquad - \alpha_B(\alpha_W + \gamma\alpha_N)(\gamma-1)(\alpha_B + \alpha_W + \gamma\alpha_N).
\end{aligned}$$

Thus, if $p(\sigma_l(\hat{h})) > 0$, then $\Delta_l^1 > 0$. To facilitate the analysis, $p(\sigma_l(\hat{h})) = a\sigma_l(\hat{h})^2 + b\sigma_l(\hat{h}) + c$ stands for (11). The first derivative is $p'(\sigma_l(\hat{h})) = 2a\sigma_l(\hat{h}) + b$ and the second one is $p''(\sigma_l(\hat{h})) = 2a$. Therefore, $p(\sigma_l(\hat{h}))$ is convex in $\sigma_l(\hat{h})$ because $a = \left(2\gamma\alpha_N^2(1-\gamma)^2(\alpha_W - \alpha_B)\right) > 0$ (it has been assumed that $\alpha_W > \alpha_B$). The first derivative gives the minimum value of $p(\sigma_l(\hat{h}))$: $p'(\sigma_l(\hat{h})_{\min}) = 2a\sigma_l(\hat{h})_{\min} + b = 0 \Leftrightarrow \sigma_l(\hat{h})_{\min} = \frac{-b}{2a}$. Consequently, $p(\sigma_l(\hat{h})_{\min}) > 0 \Rightarrow p(\sigma_l(\hat{h})) > 0 \Rightarrow \Delta_l^1 > 0$. Let us determine when $p(\sigma_l(\hat{h})_{\min}) > 0$. Note that, $p(\sigma_l(\hat{h})_{\min}) = a(\frac{-b}{2a})^2 + b(\frac{-b}{2a}) + c = c - \frac{b^2}{4a}$

From (11),

$$\frac{b^2}{4a} = \frac{\left(\alpha_N(1-\gamma)\left(-4\gamma^2\alpha_N\alpha_B + 4\gamma^2\alpha_N\alpha_W - 4\gamma\alpha_B\alpha_W + 2\gamma\alpha_N\alpha_B + 2\gamma\alpha_W^2 - 3\gamma\alpha_N\alpha_W + \alpha_B^2 - \alpha_W^2\right)\right)^2}{4\left(2\gamma\alpha_N^2(1-\gamma)^2(\alpha_W - \alpha_B)\right)}$$

$$= \frac{\left(-4\gamma^2\alpha_N\alpha_B + 4\gamma^2\alpha_N\alpha_W - 4\gamma\alpha_B\alpha_W + 2\gamma\alpha_N\alpha_B + 2\gamma\alpha_W^2 - 3\gamma\alpha_N\alpha_W + \alpha_B^2 - \alpha_W^2\right)^2}{8\gamma(\alpha_W - \alpha_B)},$$

and $c - \frac{b^2}{4a}$ is equal to

$$\Big((\gamma(\alpha_W + \gamma\alpha_N)(\alpha_B - \alpha_N(\gamma-1))(\alpha_B - \alpha_W) - \alpha_W(\gamma-1)(\alpha_B - \alpha_N(\gamma-1))(\alpha_B + \alpha_W + \gamma\alpha_N) - \alpha_B(\alpha_W + \gamma\alpha_N)(\gamma-1)(\alpha_B + \alpha_W + \gamma\alpha_N))(8\gamma(\alpha_W - \alpha_B))$$
$$- \left(-4\gamma^2\alpha_N\alpha_B + 4\gamma^2\alpha_N\alpha_W - 4\gamma\alpha_B\alpha_W + 2\gamma\alpha_N\alpha_B + 2\gamma\alpha_W^2 - 3\gamma\alpha_N\alpha_W + \alpha_B^2 - \alpha_W^2\right)^2 \Big) \frac{1}{8\gamma(\alpha_W - \alpha_B)}.$$

The denominator is positive because $\alpha_W > \alpha_B$. The sign of the numerator determines the sign of that expression. Although tedious, it is straightforward to expand and simplify the numerator of the above expression. The result is:

$8\gamma^2\alpha_B^3\alpha_W - 4\gamma^2\alpha_B^2\alpha_N^2 - 8\gamma^2\alpha_B\alpha_W^3 + 4\gamma^2\alpha_B\alpha_W\alpha_N^2 - 4\gamma^2\alpha_W^4 - 4\gamma^2\alpha_W^3\alpha_N - \gamma^2\alpha_W^2\alpha_N^2 - 8\gamma\alpha_B^3\alpha_W - 4\gamma\alpha_B^3\alpha_N$

$-4\gamma\alpha_B^2\alpha_W^2 - 2\gamma\alpha_B^2\alpha_W\alpha_N + 8\gamma\alpha_B\alpha_W^3 + 4\gamma\alpha_B\alpha_W^2\alpha_N + 4\gamma\alpha_W^4 + 2\gamma\alpha_W^3\alpha_N - \alpha_B^4 + 2\alpha_B^2\alpha_W^2 - \alpha_W^4$

It can be rewritten as a polynomial in γ:

$$pol(\gamma) = \gamma^2\left(8\alpha_B^3\alpha_W - 4\alpha_B^2\alpha_N^2 - 8\alpha_B\alpha_W^3 + 4\alpha_B\alpha_W\alpha_N^2 - 4\alpha_W^4 - 4\alpha_W^3\alpha_N - \alpha_W^2\alpha_N^2\right)$$
$$+\gamma\left(8\alpha_B\alpha_W^3 - 4\alpha_N\alpha_B^3 - 4\alpha_B^2\alpha_W^2 - 2\alpha_N\alpha_B^2\alpha_W - 8\alpha_B^3\alpha_W + 4\alpha_N\alpha_B\alpha_W^2 + 4\alpha_W^4 + 2\alpha_N\alpha_W^3\right)$$
$$+\left(2\alpha_B^2\alpha_W^2 - \alpha_B^4 - \alpha_W^4\right)$$

Simplifying the coefficients of the polynomial,

$$\begin{aligned} pol(\gamma) = {} & \gamma^2\left(8\alpha_B^3\alpha_W - 4\alpha_B^2\alpha_N^2 - 8\alpha_B\alpha_W^3 + 4\alpha_B\alpha_W\alpha_N^2 - 4\alpha_W^4 - 4\alpha_W^3\alpha_N - \alpha_W^2\alpha_N^2\right) \\ & + \gamma\left(2(\alpha_W - \alpha_B)(2\alpha_B + \alpha_W)(\alpha_B + \alpha_W)(2\alpha_W + \alpha_N)\right) \\ & - \left(\alpha_B^2 - \alpha_W^2\right)^2 \end{aligned} \tag{12}$$

Thus, $pol(\gamma) > 0 \Leftrightarrow p(\sigma_l(\hat{h})_{\min}) > 0 \Rightarrow p(\sigma_l(\hat{h})) > 0 \Leftrightarrow \Delta_l^1 > 0$. Let us determine when $pol(\gamma)$ is greater than zero.

The polynomial $pol(\gamma)$ is concave in γ because the coefficient of γ^2 is negative, as it is proved below:

$8\alpha_B^3\alpha_W + 4\alpha_B\alpha_W\alpha_N^2 - 4\alpha_W^3\alpha_N - 4\alpha_B^2\alpha_N^2 - 8\alpha_B\alpha_W^3 - 4\alpha_W^4 - \alpha_W^2\alpha_N^2$

$= \alpha_N^2\left(4\alpha_B\alpha_W - 4\alpha_B^2 - \alpha_W^2\right) - 4\alpha_W^3\alpha_N + 8\alpha_B^3\alpha_W - 8\alpha_B\alpha_W^3 - 4\alpha_W^4$

$= -\alpha_N^2\left(2\alpha_B - \alpha_W\right)^2 - 4\alpha_W^3\alpha_N + 4\alpha_W\left(2\alpha_B^3 - \alpha_W^2(2\alpha_B + 1)\right)$

$= -\alpha_N^2\left(2\alpha_B - \alpha_W\right)^2 - 4\alpha_W^3\alpha_N + 4\alpha_W\left(\alpha_B^2 2\alpha_B - \alpha_W^2(2\alpha_B + 1)\right)$

The last expression is negative if $\alpha_B^2 2\alpha_B - \alpha_W^2(2\alpha_B + 1)$ is negative, which is the case if $\alpha_W > \alpha_B$. Consequently $pol(\gamma)$ is concave if $\alpha_W > \alpha_B$.

On the other hand, the first derivative of $pol(\gamma)$ in $\gamma = \frac{1}{2}$ is negative:

$$\left.\frac{dpol(\gamma)}{d\gamma}\right|_{\gamma=\frac{1}{2}} = \left(8\alpha_B^3\alpha_W - 4\alpha_B^2\alpha_N^2 - 8\alpha_B\alpha_W^3 + 4\alpha_B\alpha_W\alpha_N^2 - 4\alpha_W^4 - 4\alpha_W^3\alpha_N - \alpha_W^2\alpha_N^2\right)$$
$$+\left(2(\alpha_W - \alpha_B)(2\alpha_B + \alpha_W)(\alpha_B + \alpha_W)(2\alpha_W + \alpha_N)\right)$$

$= -4\alpha_B^3\alpha_N - 4\alpha_B^2\alpha_W^2 - 2\alpha_B^2\alpha_W\alpha_N - 4\alpha_B^2\alpha_N^2 + 4\alpha_B\alpha_W^2\alpha_N + 4\alpha_B\alpha_W\alpha_N^2 - 2\alpha_W^3\alpha_N - \alpha_W^2\alpha_N^2$

$= -4\alpha_B^2\alpha_W^2 - 2\left(2\alpha_B^3 + \alpha_B^2\alpha_W - 2\alpha_B\alpha_W^2 + \alpha_W^3\right)\alpha_N - \left(-2\alpha_B + \alpha_W\right)^2\alpha_N^2$
$= -4\alpha_B^2\alpha_W^2 - 2\left(2\alpha_B^3 + \alpha_W\left(\alpha_B^2 - 2\alpha_B\alpha_W + \alpha_W^2\right)\right)\alpha_N - \left(-2\alpha_B + \alpha_W\right)^2\alpha_N^2$
$= -4\alpha_B^2\alpha_W^2 - 2\left(2\alpha_B^3 + \alpha_W\left(\alpha_B - \alpha_W\right)^2\right)\alpha_N - \left(-2\alpha_B + \alpha_W\right)^2\alpha_N^2 > 0$

Clearly, the last expression is greater than zero and consequently $\left.\frac{dpol(\gamma)}{d\gamma}\right|_{\gamma=\frac{1}{2}} < 0$. Thus, as $pol(\gamma)$ is concave for $\gamma \in (\frac{1}{2}, 1)$ and decreasing in $\gamma = \frac{1}{2}$, polynomial $pol(\gamma)$ is necessarily decreasing from $\gamma = \frac{1}{2}$ to $\gamma = 1$. Therefore, there are only two possibilities. In the first one, $pol(\gamma)$ is negative for $\gamma \in \left(\frac{1}{2}, 1\right)$. The second one is that $pol(\gamma)$ is positive for $\gamma \in (\frac{1}{2}, \tilde{\gamma})$ and negative if $\gamma > \tilde{\gamma}$, where $\tilde{\gamma}$ is the greatest real root of $pol(\gamma) = 0$ (note that $pol(\gamma)$ is a second-degree polynomial). If $pol(\gamma = \frac{1}{2}) > 0$, we are in the second case.

With $\alpha_N = (1 - \alpha_B - \alpha_W)$, from equation (12) it is straightforward to derive that,

$pol(\gamma = \frac{1}{2}) = -\frac{1}{4}\alpha_B^2\alpha_W^2 - \alpha_B^2\alpha_W - \alpha_B^2 + \frac{1}{2}\alpha_B\alpha_W^3 + \frac{1}{2}\alpha_B\alpha_W^2 + \alpha_B\alpha_W - \frac{1}{4}\alpha_W^4 + \frac{1}{2}\alpha_W^3 - \frac{1}{4}\alpha_W^2$
$= \left(-\frac{1}{4}\alpha_W^2 - \alpha_W - 1\right)\alpha_B^2 + \left(\frac{1}{2}\alpha_W^3 + \frac{1}{2}\alpha_W^2 + \alpha_W\right)\alpha_B + \left(\frac{1}{2}\alpha_W^3 - \frac{1}{4}\alpha_W^4 - \frac{1}{4}\alpha_W^2\right).$

The expression above is concave in α_B, and it has the followings real roots:
$\left\{\frac{\left(2\alpha_W + \alpha_W^2 + \alpha_W^3 - 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}\right)}{(\alpha_W+2)^2}, \frac{\left(2\alpha_W + \alpha_W^2 + \alpha_W^3 + 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}\right)}{(\alpha_W+2)^2}\right\}.$

Consequently, if $\alpha_B \in \left(\frac{\left(2\alpha_W + \alpha_W^2 + \alpha_W^3 - 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}\right)}{(\alpha_W+2)^2}, \frac{\left(2\alpha_W + \alpha_W^2 + \alpha_W^3 + 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}\right)}{(\alpha_W+2)^2}\right)$, the $pol(\gamma = \frac{1}{2}) > 0$ and $pol(\gamma)$ is positive for $\gamma \in (\frac{1}{2}, \tilde{\gamma})$ and negative if $\gamma > \tilde{\gamma}$, where $\tilde{\gamma}$ has to be the greatest real root of $pol(\gamma)$. With $\alpha_N = (1 - \alpha_B - \alpha_W)$, the two roots of $pol(\gamma)$ are $\frac{\left(\alpha_W^2 - \alpha_B^2\right)}{(2\alpha_B+\alpha_W)(1-\alpha_B+\alpha_W)+2\sqrt{2\alpha_B\alpha_W}} < \frac{\left(\alpha_W^2 - \alpha_B^2\right)}{(2\alpha_B+\alpha_W)(1-\alpha_B+\alpha_W)-2\sqrt{2\alpha_B\alpha_W}}$, therefore, $\tilde{\gamma} = \frac{\left(\alpha_B^2 - \alpha_W^2\right)}{-2\sqrt{2\alpha_B\alpha_W}+(2\alpha_B+\alpha_W)(1-\alpha_B+\alpha_W)}$.

The following claim is necessary to complete the proof.

Claim 14. *For any α_W,*
$\frac{\left(\alpha_B^2 - \alpha_W^2\right)}{2\sqrt{2\alpha_B\alpha_W}-(2\alpha_B+\alpha_W)(1-\alpha_B+\alpha_W)} < \alpha_B^{**} = \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right) < \frac{2\alpha_W + \alpha_W^2 + \alpha_W^3 + 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}}{(\alpha_W+2)^2}.$

Proof

First we prove that,

$\frac{2\alpha_W + \alpha_W^2 + \alpha_W^3 + 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}}{(\alpha_W+2)^2} > \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$
$\iff \frac{1}{4}\left(\frac{8\alpha_W + 4\alpha_W^2 + 4\alpha_W^3 + 8\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3} - (\alpha_W+2)^2(2+\alpha_W) + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4}}{(\alpha_W+2)^2}\right) > 0$
$\iff 8\alpha_W + 4\alpha_W^2 + 4\alpha_W^3 - (\alpha_W+2)^2(2+\alpha_W) + 8\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3} + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4} > 0$
$\iff 3\alpha_W^3 + 8\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3} - 2\alpha_W^2 - 4\alpha_W - 8 + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4} > 0$
$\Longleftarrow -2\alpha_W^2 - 4\alpha_W - 8 + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4} > 0$
$\iff \sqrt{9\alpha_W^2 - 4\alpha_W + 4} > \frac{2\alpha_W^2 + 4\alpha_W + 8}{(\alpha_W+2)^2}$
$\iff 9\alpha_W^2 - 4\alpha_W + 4 - \left(\frac{2\alpha_W^2 + 4\alpha_W + 8}{(\alpha_W+2)^2}\right)^2 > 0$
$\iff \frac{9\alpha_W^6 + 68\alpha_W^5 + 184\alpha_W^4 + 208\alpha_W^3 + 64\alpha_W^2}{(\alpha_W+2)^4} > 0.$

Now, we prove that

$\frac{2\alpha_W + \alpha_W^2 + \alpha_W^3 - 2\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3}}{(\alpha_W+2)^2} < \frac{1}{4}\left(2 + \alpha_W - \sqrt{9\alpha_W^2 - 4\alpha_W + 4}\right)$
$\iff \frac{1}{4}\left(\frac{8\alpha_W + 4\alpha_W^2 + 4\alpha_W^3 - 8\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3} - (\alpha_W+2)^2(2+\alpha_W) + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4}}{(\alpha_W+2)^2}\right) < 0$
$\iff 8\alpha_W + 4\alpha_W^2 + 4\alpha_W^3 - (\alpha_W+2)^2(2+\alpha_W) - 8\sqrt{2}\sqrt{\alpha_W^4 + \alpha_W^3} + (\alpha_W+2)^2\sqrt{9\alpha_W^2 - 4\alpha_W + 4} < 0$
$\iff 3\alpha_W^3 - 2\alpha_W^2 - 4\alpha_W - 8 - \sqrt{\left(8\sqrt{2}\right)^2\left(\alpha_W^4 + \alpha_W^3\right)} + \sqrt{(\alpha_W+2)\left(9\alpha_W^2 - 4\alpha_W + 4\right)} < 0$
$\iff 3\alpha_W^3 - 2\alpha_W^2 - 4\alpha_W - 8 - \sqrt{128\alpha_W^4 + 128\alpha_W^3} + \sqrt{9\alpha_W^3 + 14\alpha_W^2 - 4\alpha_W + 8} < 0$

$\Longleftarrow 3\alpha_W^3 - 2\alpha_W^2 - 4\alpha_W - 8 - \sqrt{128\alpha_W^4 + 128\alpha_W^3} + \left(9\alpha_W^3 + 14\alpha_W^2 - 4\alpha_W + 8\right) < 0$
$\Longleftrightarrow 3\alpha_W^3 - 2\alpha_W^2 - 4\alpha_W - 8 + 9\alpha_W^3 + 14\alpha_W^2 - 4\alpha_W + 8 < \sqrt{128\alpha_W^4 + 128\alpha_W^3}$
$\Longleftrightarrow 4\alpha_W\left(3\alpha_W + 3\alpha_W^2 - 2\right) < \sqrt{128\alpha_W^4 + 128\alpha_W^3}$
$\Longleftrightarrow 3\alpha_W + 3\alpha_W^2 - 2 < \sqrt{\frac{128\alpha_W^4 + 128\alpha_W^3}{(4\alpha_W)^2}}$
$\Longleftrightarrow 3\alpha_W + 3\alpha_W^2 - 2 < \sqrt{8\alpha_W^2 + 8\alpha_W}$

It is straightforward to show that the left side of the last inequality is convex in α_W and the right side is concave. In addition, the left side is lower than the right side in $\alpha_W = 0$, and in $\alpha_W = 1$ they are equal. Therefore, the inequality holds. ⧫

Therefore, with $\mu = 1$, if $\alpha_B \in (\alpha_B^*, \alpha_B^{**})$ and $\gamma \in (\frac{1}{2}, \tilde{\gamma})$, then $pol(\gamma) > 0$, which implies that $p(\sigma_l(\hat{h})) > 0$, which in turn implies that $\Delta_l^1 > 0$. Consequently, in equilibrium always $\sigma_l(\hat{h})^* = 1$. ■

Proof of Proposition 5

We first define $\sigma_l^\mu(\hat{l})^{Sup*}$ to be the supreme of all possible equilibrium strategy after signal l, $\sigma_l^\mu(\hat{l})^*$.

By Proposition 1, $\sigma_h^\mu(\hat{h})^* = 1$. Hereafter, we write $\Delta_l^\mu[\sigma_l(\hat{l}), \sigma_h(\hat{h}) = 1]$ as $\Delta_l^\mu[\sigma_l(\hat{l})]$. If $\alpha_B \in (\alpha_B^*, \alpha_B^{**})$ and $\gamma \in (\frac{1}{2}, \tilde{\gamma})$, then $\Delta_l^1[\sigma_l(\hat{l})] > 0$ for all $\sigma_l(\hat{l}) \in (0,1)$. In addition $\Delta_l^0[\sigma_l(\hat{l})] < 0$ for all $\sigma_l(\hat{l}) \in \left(0, \sigma_l^0(\hat{l})^*\right)$, and $\Delta_l^0[\sigma_l(\hat{l})] > 0$ for all $\sigma_l(\hat{l}) \in \left(\sigma_l^0(\hat{l})^*, 1\right)$ (see Proposition 4). Therefore, $\Delta_l^\mu[\sigma_l(\hat{l})] = (1-\mu)\Delta_l^0[\sigma_l(\hat{l})] + \mu\Delta_l^1[\sigma_l(\hat{l})] > 0$ for all $\sigma_l(\hat{l}) \in \left(\sigma_l^0(\hat{l})^*, 1\right)$, which implies that $\sigma_l^\mu(\hat{l})^{Sup*} < \sigma_l^{\mu=0}(\hat{l})^*$ for any $\mu > 0$. Thus, $\sigma_l^\mu(\hat{l})^{Sup*} \in \left(0, \sigma_l^0(\hat{l})^*\right)$.

Note that $\frac{\partial \Delta_l^\mu[\sigma_l(\hat{l})]}{\partial \mu} > 0$ for all $\sigma_l(\hat{l}) \in \left(0, \sigma_l^0(\hat{l})^*\right)$, because $\Delta_l^1[\sigma_l(\hat{l})] > 0$ for all $\sigma_l(\hat{l}) \in (0,1)$ and $\Delta_l^{\mu=0}[\sigma_l(\hat{l})] < 0$ for all $\sigma_l(\hat{l}) \in \left(0, \sigma_l^0(\hat{l})^*\right)$.

Therefore, if μ increases, the function $\Delta_l^\mu[\sigma_l(\hat{l})]$ increases and, as $\Delta_l^\mu[\sigma_l(\hat{l})] > 0$ for all $\sigma_l(\hat{l}) \in \left(0, \sigma_l^\mu(\hat{l})^{Sup*}\right)$, $\sigma_l^\mu(\hat{l})^{Sup*}$ has to decrease with μ. ■

Remark 2. *Let $u(\theta)$ be the utility the judge obtains when the principal believes the judge to be of type θ. Suppose $u(W) > u(N) > u(B) = 0$, i.e. we normalize the utility of being considered a biased type to zero. With a risk-neutral judge, her expected payoff is:*

$\lambda_W[l,X]u(W) + \lambda_N[l,X]u(N) + \lambda_B[l,X]u(B)$
$= \lambda_W[l,X]u(W) + \lambda_N[l,X]u(N)$
$= \lambda_W[l,X]u(W) + (1 - \lambda_W[l,X] - \lambda_B[l,X])u(N)$
$= \lambda_W[l,X]\left(u(W) - u(N)\right) + (1 - \lambda_B[l,X])u(N)$
$= \lambda_W[l,X]\left(u(W) - u(N)\right) + \lambda_{\bar{B}}[l,X]u(N).$

Now, let $\vartheta = (u(W) - u(N))$ and $\beta = u(N)$.

Substituting, the expression above is: $\vartheta\lambda_W[l,X] + \beta\lambda_{\bar{B}}[l,X]$.

That is, under the assumption that the judge is risk-neutral, we can write the more general judge's objective function $f(\cdot)$, as defined by equation (1), as a linear function. Now, if we assume $\vartheta = \beta$, we have the objective function defined in equation (8).

B Appendix

We now analyze a more general game in which both, the wise judge and the normal judge, are strategic. However, they are different in terms of their information structure. As in the main body of the paper, we assume that the wise judge receives a perfectly informative signal of quality 1, and that the normal judges receives a signal of quality $\gamma > 1/2$.

We introduce some notation to refer to the wise judge. Notation for the normal judge is the same than in the main body of the paper. Thus, a profile of strategies is denoted by $(\sigma_N, \sigma_W) = \Big((\sigma_l(\hat{l}), \sigma_h(\hat{h})), (\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h}))\Big)$, where the first strategy stands for the normal type and the second for the wise type. It is straightforward to derive the analogous of equations (9) and (10) for the wise type, which are denoted by $\Delta^{\mu}_{W,l}[\cdot]$ and $\Delta^{\mu}_{W,h}[\cdot]$ (see equations (13) and (14) below). Note that equations (13) and (14) are analogous to equations (9) and (10), with $\gamma = 1$. Also, note that equations (9)-(10) and (13)-(14) depend on the strategy profile. We make it explicit when needed. Last, note that beliefs in Tables 1 and 2 were obtained under the assumption that the wise judge follows her signal. Here, we specify the new beliefs when it is necessary.

$$\begin{aligned}\Delta^{\mu}_{W,l}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] &= (1-\mu)\Delta^{0}_{W,l}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] + \mu\Delta^{1}_{W,l}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] \\ &= (1-\mu)\left(f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0], \lambda_{\bar{B}}[\hat{l},0]\right)\right) + \\ &\mu\left(f(0,1) - f\left(\lambda_W[\hat{l},L], \lambda_{\bar{B}}[\hat{l},L]\right)\right)\end{aligned} \tag{13}$$

$$\begin{aligned}\Delta^{\mu}_{W,h}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] &= (1-\mu)\Delta^{0}_{W,h}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] + \mu\Delta^{1}_{W,h}[\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})] \\ &= (1-\mu)\left(f\left(\lambda_W[\hat{h},0],1\right) - f\left(\lambda_W[\hat{l},0], \lambda_{\bar{B}}[\hat{l},0]\right)\right) + \\ &\mu\left(f\left(\lambda_W[\hat{h},H],1\right) - f\left(0, \lambda_{\bar{B}}[\hat{l},H]\right)\right)\end{aligned} \tag{14}$$

Next proposition shows that the honest strategy, $\bar{\sigma}_W = \Big(\bar{\sigma}_{W,l}(\hat{l}) = 1, \bar{\sigma}_{W,h}(\hat{h}) = 1\Big)$, is the unique equilibrium strategy for the wise type in any equilibrium in which the normal type takes action $\hat{l}$ with positive probability, i.e. $\sigma^{\mu}_{l}(\hat{l})^* > 0$. Proposition 7 analyzes the case in which $\sigma^{\mu}_{l}(\hat{l})^*$ is equal to zero. In that case, as a consequence of Lemma 8, we have $\sigma^{\mu}_{h}(\hat{h})^* = 1$. That is, the normal type always takes action $\hat{h}$, irrespectively of her signal. Note that this is a less interesting scenario and that in this case, media coverage can never have a perverse effect.

Proposition 6. *For any increasing function $f(\lambda_W(a,X), \lambda_{\bar{B}}(a,X))$, any $\mu \in (0,1)$, and any equilibrium in which the normal type makes $\sigma^{\mu}_{l}(\hat{l})^* > 0$, the honest strategy, $\bar{\sigma}_W = \Big(\bar{\sigma}_{W,l}(\hat{l}) = 1, \bar{\sigma}_{W,h}(\hat{h}) = 1\Big)$, is the unique equilibrium strategy for the wise type, $\sigma^*_W = \bar{\sigma}_W$.*

Proof

First we prove the following two lemmas.

Lemma 7. *In equilibrium, $\lambda_{\bar{B}}[\hat{l},H] < \lambda_{\bar{B}}[\hat{l},L]$*

Proof

When the wise type can play any strategy, $\Big(\sigma_{W,l}(\hat{l}), \sigma_{W,h}(\hat{h})\Big)$, the beliefs are

$\lambda_B(\hat{l},L) = \frac{P(\hat{l}|B,L)P(L)P(B)}{P(\hat{l}|B,L)P(L)P(B)+P(\hat{l}|N,L)P(L)P(N)+P(\hat{l}|W,L)P(L)P(W)} = \frac{\alpha_B}{\alpha_B+\left(\gamma\sigma_l(\hat{l})+(1-\gamma)\sigma_h(\hat{l})\right)\alpha_N+\sigma_{W,l}(\hat{l})\alpha_W}$,

$\lambda_B(\hat{l},H) = \frac{P(\hat{l}|B,H)P(H)P(B)}{P(\hat{l}|B,H)P(H)P(B)+P(\hat{l}|N,H)P(H)P(N)+P(\hat{l}|W,H)P(H)P(W)} = \frac{\alpha_B}{\alpha_B+\left((1-\gamma)\sigma_l(\hat{l})+\gamma\sigma_h(\hat{l})\right)\alpha_N+\sigma_{W,h}(\hat{l})\alpha_W}$,

where,

$\frac{\lambda_B(\hat{l},L)}{\lambda_B(\hat{l},H)} > 1 \iff \frac{\big((1-\gamma)\sigma_l(\hat{l})+\gamma\sigma_h(\hat{l})\big)\alpha_N+\sigma_{W,h}(\hat{l})\alpha_W}{\big(\gamma\sigma_l(\hat{l})+(1-\gamma)\sigma_h(\hat{l})\big)\alpha_N+\sigma_{W,l}(\hat{l})\alpha_W} > 1,$
which holds for the following two reasons:

On the one hand, $(1-\gamma)\sigma_l(\hat{l})+\gamma\sigma_h(\hat{l}) > \gamma\sigma_l(\hat{l})+(1-\gamma)\sigma_h(\hat{l}) \iff (1-2\gamma)\sigma_l(\hat{l}) > (1-2\gamma)\sigma_h(\hat{l}) \iff \sigma_l(\hat{l}) < \sigma_h(\hat{l})$, which is always the case in a non perverse equilibrium.

On the other hand, $\sigma_{W,l}(\hat{l}) < \sigma_{W,h}(\hat{l})$ for the same reason.

Therefore, $\lambda_B(\hat{l},L) > \lambda_B(\hat{l},H)$, and consequently, $\lambda_{\bar{B}}[\hat{l},H] < \lambda_{\bar{B}}[\hat{l},L]$. ♦

Lemma 8. *For any increasing function $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X))$ and any $\mu \in (0,1)$, it always holds that in equilibrium $\Delta^\mu_{W,l}[\sigma_W,\sigma_N] < \Delta^\mu_l[\sigma_W,\sigma_N] < \Delta^\mu_h[\sigma_W,\sigma_N] < \Delta^\mu_{W,h}[\sigma_W,\sigma_N]$.*

Proof

From equations (9)-(10), (13)-(14) and Lemma 7, it is straightforward to prove that

$\Delta^\mu_{W,l}[\sigma_W,\sigma_N] < \Delta^\mu_l[\sigma_W,\sigma_N] \iff f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right) > f\left(0,\lambda_{\bar{B}}[\hat{l},H]\right)$,

$\Delta^\mu_l[\sigma_W,\sigma_N] < \Delta^\mu_h[\sigma_W,\sigma_N] \iff f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right) > f\left(0,\lambda_{\bar{B}}[\hat{l},H]\right)$,

$\Delta^\mu_h[\sigma_W,\sigma_N] < \Delta^\mu_{W,h}[\sigma_W,\sigma_N] \iff f\left(\lambda_W[\hat{l},L],\lambda_{\bar{B}}[\hat{l},L]\right) > f\left(0,\lambda_{\bar{B}}[\hat{l},H]\right)$. By Lemma 7 these expressions hold. ♦

First, we prove by contradiction that $\sigma^\mu_{W,h}(\hat{h})^* = 1$. Let's assume that $\sigma^\mu_{W,h}(\hat{h})^* < 1$. It implies that $\Delta^\mu_{W,h}[\sigma^*_W,\sigma^*_N] \le 0$. By Lemma 8, $\Delta^\mu_{W,l}[\sigma^*_W,\sigma^*_N] < \Delta^\mu_l[\sigma^*_W,\sigma^*_N] < \Delta^\mu_h[\sigma^*_W,\sigma^*_N] < \Delta^\mu_{W,h}[\sigma^*_W,\sigma^*_N] \le 0$. Consequently $\sigma^\mu_{W,l}(\hat{l})^* = 1$, $\sigma^\mu_l(\hat{l})^* = 1$ and $\sigma^\mu_h(\hat{h})^* = 0$. This implies that if the wise type takes action $\hat{h}$, the equilibrium beliefs assign probability one to the judge being wise and probability zero to her being a biased type (maximum possible payoff). Thus, this payoff is always greater than if action $\hat{l}$ is taken. Therefore $\sigma^\mu_{W,h}(\hat{h})^*$ cannot be smaller than one. Consequently, $\sigma^\mu_{W,h}(\hat{h})^* = 1$.

Second, we prove that $\sigma^\mu_{W,l}(\hat{l})^* = 1$ if $\sigma^\mu_l(\hat{l})^* > 0$. Note that, $\Delta^\mu_l[\sigma^*_W,\sigma^*_N] \le 0$ if $\sigma^\mu_l(\hat{l})^* > 0$. Thus, $\Delta^\mu_{W,l}[\sigma^*_W,\sigma^*_N] < \Delta^\mu_l[\sigma^*_W,\sigma^*_N] \le 0$, which implies that $\sigma^\mu_{W,l}(\hat{l})^* = 1$. ■

Next, we analyze the case in which $\sigma^\mu_l(\hat{l})^* = 0$. That is, we consider an equilibrium in which the normal type never takes action $\hat{l}$ after signal l. As we will show in the next proposition, if $\sigma^\mu_l(\hat{l})^* = 0$, then necessarily $\sigma^\mu_h(\hat{h})^* = 1$ and $\sigma^\mu_{W,h}(\hat{h})^* = 1$. That is, in this case the normal type always takes action $\hat{h}$, irrespective of her signal, and the wise judge always passes harsh sentences after signal h. Consequently, a principal that observes action $\hat{l}$ must assign zero probability to the judge being a normal type. As for the wise judge, whether she take action $\hat{l}$ or $\hat{h}$ after signal l depends on the prior probability that a judge is bias, i.e. α_B. In order to analyze the wise judge's behavior in the case $\sigma^\mu_l(\hat{l})^* = 0$, we take $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X))$ to be linear, and we prove that the honest strategy is an equilibrium strategy if $\alpha_B < \alpha^{**}_B$. Thus, in the following proposition we use $f(\lambda_W(a,X),\lambda_{\bar{B}}(a,X)) = \lambda_W(a,X)+\lambda_{\bar{B}}(a,X)$.

Proposition 7. *In any equilibrium in which the normal type makes $\sigma^\mu_l(\hat{l})^* = 0$, the honest strategy for the wise type, $\bar{\sigma}_W = \left(\bar{\sigma}_{W,l}(\hat{l}) = 1, \bar{\sigma}_{W,h}(\hat{h}) = 1\right)$, is an equilibrium strategy if $\alpha_B < \alpha^{**}_B = \frac{1}{4}\left(2+\alpha_W-\sqrt{9\alpha_W^2-4\alpha_W+4}\right)$.*

Proof

Note that if $\sigma^\mu_l(\hat{l})^* = 0$, then $\Delta^\mu_l[\sigma^*_W,\sigma^*_N] \ge 0$; and by Lemma 8, $\Delta^\mu_{W,l}[\sigma^*_W,\sigma^*_N] < \Delta^\mu_l[\sigma^*_W,\sigma^*_N] < \Delta^\mu_h[\sigma^*_W,\sigma^*_N] < \Delta^\mu_{W,h}[\sigma^*_W,\sigma^*_N]$.

Taking all together, we have $\sigma_h^{\mu}(\hat{h})^* = 1$ and $\sigma_{W,h}^{\mu}(\hat{h})^* = 1$. Then, we only have to prove that $\Delta_{W,l}^{\mu}[\sigma_W^* = \bar{\sigma}_W, \sigma_N^* = \left(\sigma_l^{\mu}(\hat{l})^* = 0, \sigma_h^{\mu}(\hat{h})^* = 1\right)] < 0$ where, in this case, the beliefs in equation (13) are those from Tables 1 and 2. With this purpose, we first prove the followings two claims.

Claim 15. $\Delta_{W,l}^{0}[\bar{\sigma}_W, \sigma_N = \left(\sigma_l^{\mu}(\hat{l})^* = 0, \sigma_h^{\mu}(\hat{h}) = 1\right)] < 0$ *if* $\alpha_B < \alpha_B^{**}$

Proof

Note that $\Delta_{W,l}^{0}[\bar{\sigma}_W, \sigma_N] = \Delta_l^{0}[\bar{\sigma}_W, \sigma_N]$. Thus, this claim is a consequence of Proposition 4. In Proposition 4 it is shown that if $\alpha_B < \alpha_B^{**}$, then $\sigma_l^{\mu}(\hat{l}) = 0$ cannot be part of an equilibrium, which means that $\Delta_l^{0}[\bar{\sigma}_W, \sigma_N] < 0$. It implies $\Delta_{W,l}^{0}[\bar{\sigma}_W, \sigma_N = \left(\sigma_l^{\mu}(\hat{l})^* = 0, \sigma_h^{\mu}(\hat{h})^* = 1\right)] < 0$. ♦

Claim 16. $\Delta_{W,l}^{1}[\bar{\sigma}_W, \sigma_N] < 0 \iff \alpha_B < \alpha_B^{**}$.

Proof

From equation (13), $\Delta_{W,l}^{1}[\bar{\sigma}_W, \sigma_N] = f(0,1) - f\left(\lambda_W[\hat{l}, L], \lambda_{\bar{B}}[\hat{l}, L]\right) = 1 - \lambda_W[\hat{l}, L] - (1 - \lambda_B[\hat{l}, L]) = \lambda_B[\hat{l}, L] - \lambda_W[\hat{l}, L]$.

From Tables 1 and 2,

$\lambda_W[\hat{l}, L] = \frac{\alpha_W}{\alpha_B + \alpha_W + ((1-\gamma)\sigma_h(\hat{l}) + \gamma\sigma_l(l))\alpha_N}$,

$\lambda_B[\hat{l}, L] = \frac{\alpha_B}{\alpha_B + \alpha_W + ((1-\gamma)\sigma_h(\hat{l}) + \gamma\sigma_l(\hat{l}))\alpha_N}$.

Therefore,

$\lambda_W[\hat{l}, L] > \lambda_B[\hat{l}, L]$ if and only if $\alpha_B < \alpha_W$, and by Claim 12, $\alpha_B^{**} < \alpha_W$. ♦

Note that $\Delta_{W,l}^{\mu}[\bar{\sigma}_W, \sigma_N] = (1-\mu)\left(\Delta_{W,l}^{0}[\bar{\sigma}_W, \sigma_N]\right) + \mu\left(\Delta_{W,l}^{1}[\bar{\sigma}_W, \sigma_N\right)$. Consequently, $\Delta_{W,l}^{\mu}[\bar{\sigma}_W, \sigma_N = \left(\sigma_l^{\mu}(\hat{l})^* = 0, \sigma_h^{\mu}(\hat{h}) = 1\right)] < 0$ if $\alpha_B < \alpha_B^{**}$. ■

References

Andina-Díaz, Ascensión and José A. García-Martínez (2015), A theory of media self-silence. METC Working Paper 4.

Banerjee, Abhijit V. (1992), 'A simple model of herd behavior', *Quarterly Journal of Economics* **107**(3), 797–817.

Bourjade, Sylvain and Bruno Jullien (2011), 'The roles of reputation and transparency on the behavior of biased experts', *RAND Journal of Economics* **42**(3), 575–594.

Caldarone, Richard P., Brandice Canes-Wrone and Tom S. Clark (2009), 'Partisan labels and democratic accontability: An analysis of State Supreme Court abortion decisions', *Journal of Politics* **71**(2), 560–573.

Canes-Wrone, Brandice, Tom S. Clark and Jee-Kwand Park (2012), 'Judicial independence and retention elections', *Journal of Law, Economics and Organization* **28**(2), 211–234.

Cohen, Alma, Alon Klement and Zvika Neeman (2015), 'Judicial decision making: A dynamic reputation approach', *Journal of Legal Studies* **44**(1), 133–159.

Cummins, Jason G. and Ingmar Nyman (2005), 'The dark side of competitive pressure', *RAND Journal of Economics* **36**(2), 361–377.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes and Andrei Shleifer (2003), 'Courts', *Quarterly Journal of Economics* **118**(2), 453–517.

Fox, Justin and Richard Van Weelden (2012), 'Costly transparency', *Journal of Public Economics* **96**(1), 142–150.

Garoupa, Nuno and Tom Ginsburg (2011), 'Hybrid judicial career structures: Reputation vs. legal tradition', *Journal of Legal Analysis* **3**(2), 411–448.

Glaeser, Edward L. and Andrei Shleifer (2002), 'Legal origins', *Quarterly Journal of Economics* **117**(4), 1193–1229.

Gordon, Sanford C. and Gregory A. Huber (2007), 'The effect of electoral competitiveness on incumbent behavior', *Quarterly Journal of Political Science* **2**(2), 107–138.

Heidhues, Paul and Johan Lagerlöf (2003), 'Hiding information in electoral competition', *Games and Economic Behavior* **42**(1), 48–74.

Huber, Gregory A. and Sanford C. Gordon (2004), 'Accountability and coercion: Is justice blind when it runs for office?', *American Journal of Political Science* **48**(2), 247–263.

Iossa, Elisabetta and Bruno Jullien (2012), 'The market for lawyers and quality layers in legal services', *RAND Journal of Economics* **43**(4), 677704.

Levy, Gilat (2005), 'Careerist judges and the appeals process', *RAND Journal of Economics* **36**(2), 275–297.

Lim, Claire S. H. (2015), 'Media influence on courts: Evidence from civil case adjudication', *American Law and Economics Review* **17**(1), 87–126.

Lim, Claire S. H., James M. Snyder and David Strömberg (2015), 'The judge, the politician, and the press: Newspaper coverage and criminal sentencing across electoral systems', *American Economic Journal: Applied Economics* **7**(4), 103–135.

Miceli, Thomas J. and Metin M. Cosgel (1994), 'Reputation and judicial decision making', *Journal of Economic Behavior and Organization* **23**(1), 31–5.

Morris, Stephen (2001), 'Political correctness', *Journal of Political Economy* **109**(2), 231–265.

Ottaviani, Marco and Peter Norman Sorensen (2006), 'Professional advice', *Journal of Economic Theory* **126**(1), 120 – 142.

Park, Kyung H. (2014), Judicial elections and discrimination in criminal sentencing. Working paper.

Prat, Andrea (2005), 'The wrong kind of transparency', *American Economic Review* **95**(3), 862–877.

Rubin, Paul H. and Joanna M. Shepherd (2013), *The Elgar Companion to Public Choice*, Edward Elgar, chapter Public Choice and the Law, pp. 345–362.

Scharfstein, David S. and Jeremy C. Stein (1990), 'Herd behavior and investment', *American Economic Review* **80**(3), 465–479.