Wrongful Conviction, Persuasion and Loss Aversion

Matthew J. Robertson

December 2018 No: 48

Wrongful Conviction, Persuasion and Loss Aversion

Matthew J. Robertson[*][†]

January 8, 2019

Abstract

When can a prosecutor persuade a loss-averse judge to increase her rate of conviction? Motivated by empirical evidence, I study a model of persuasion in which the loss a judge incurs from wrongful conviction looms larger than the gain from a just verdict. I show that, surprisingly, the prosecutor benefits from persuasion even when the judge is extremely loss-averse. However, a necessary condition is that the prosecutor does not underestimate the judge's loss aversion. I draw on experimental findings to quantify the effectiveness of persuasion under loss aversion.

Key Words: information design; loss aversion; wrongful conviction; Bayesian persuasion.

JEL Codes: D72; D82; D91; K40.

1 Introduction

Through persuasion, a prosecutor can increase a judge's probability of conviction despite sharing a common prior over the defendant's guilt (Kamenica and Gentzkow, 2011). However, it assumes that the judge receives an equivalent payoff from convicting the innocent and acquitting the guilty. In reality, evidence suggests that losses from wrongful conviction loom larger.

Wrongful convictions have serious consequences for both judges and the economy. The challenges faced by judges have been characterised as "mark[ing] the way for a plague of followers that deplete trials of fairness" (Traynor, 1970) and a decrease in criminal deterrence (Garoupa and Rizzolli, 2012). Moreover, the economic implications are substantial, costing taxpayers $282m over the last twenty-four years in California alone (Silbert et al., 2015). Therefore, understanding the effectiveness of persuasion when wrongful conviction carries negative consequences is an important question in information design.

In this paper, I model the judge's asymmetric payoff structure with loss-averse preferences. Specifically, the negative payoff the judge receives from wrongful conviction looms larger than the positive payoff received from a just verdict.

[*] Thanks are due to Jasmin Droege and Marco Fongoni for insightful comments.

[†] Coventry University, Department of Economics, School of Economics, Finance and Accounting, William Morris Building, Gosford Street, Coventry, CV1 5DL. Email: ac9969@coventry.ac.uk

One would expect that, as the judge becomes severely loss-averse, she is increasingly dubious of signals of guilt sent by the prosecutor. Therefore, intuition would suggest that, after a certain threshold, the judge would be sufficiently sceptical of guilty signals that the prosecutor would cease to gain from persuasion.

I show that, surprisingly, the prosecutor always gains from persuasion, irrespective of the extent of the judge's loss aversion. Compared to a fully informative investigation, the prosecutor's benefit from persuasion is strict when the judge is finitely loss-averse. However, as the judge becomes more loss-averse, persuasion is less effective. As loss aversion approaches infinity, the judge's conviction rate approaches the prior. A necessary condition for the prosecutor to optimally gain from persuasion is common knowledge of the degree of the judge's loss aversion. If the prosecutor underestimates the judge's loss aversion, persuasion becomes impossible.

I build on the literature that considers the interplay between behavioural mechanisms and information design. For example, when agents are present-biased, traffic-light information nudges are optimal (Mariotti et al., 2018). Loss aversion affects whether information should be revealed sequentially or simultaneously to consumers (Liu, 2018) and, in a dynamic setting, either full information disclosure or no communication is optimal (Lipnowski and Mathevet, 2018).

My contribution is the insight that the prosecutor benefits from persuasion even when the judge is highly loss-averse. I show this by generalising Kamenica and Gentzkow's (2011) motivating example, which has previously been extended to include costly signals for the prosecutor (Gentzkow and Kamenica, 2014). A similar example has analysed the implications of costly information acquisition for the sender (Matysková, 2018)[1]. A further contribution is demonstrating that this insight rests upon the prosecutor's precise knowledge of the judge's loss aversion.

In Section 2, I set out the model and, in Section 3, I state my main results. In Section 4, I draw on experimental evidence to quantify the reduction in the effectiveness of persuasion. Finally, I conclude in Section 5 and suggest avenues for future research. Proofs are in Appendix A.

2 Model

I extend Kamenica and Gentzkow's (2011) example to include a loss-averse judge and general payoffs. The defendant is innocent or guilty $\omega \in \Omega := \{I, G\}$ and the judge's action $a \in \mathcal{A} := \{A, C\}$ is to acquit or convict. The judge's payoffs are

$$u(a,\omega) = \begin{cases} \Pi_J & \text{if } (C,G) \text{ or } (A,I) \\ 0 & \text{if } (A,G) \\ -\lambda & \text{if } (C,I) \end{cases}$$

where $\Pi_J > 0$ and, to capture loss aversion, $\lambda > \Pi_J$. The prosecutor's payoffs are

$$v(a) = \begin{cases} \Pi_P & \text{if } a = C \\ 0 & \text{if } a = A \end{cases}$$

[1] Gentzkow and Kamenica (2014) and Matysková (2018) also provide general analysis of these extensions.

where $\Pi_P > 0$. The common prior over the defendant's guilt is $\Pr(\omega = G) := \mu_0 \in (0,1)$. The prosecutor chooses an investigation $\pi(s|\omega)$ to determine the probability the judge receives recommendation $s \in \{i, g\}$ given state ω. Under posterior belief μ_s, induced by signal realisation s, the judge's optimal action is

$$a^*(\mu_s) := \underset{a \in \mathcal{A}}{\arg\max}\ \mathbb{E}_{\mu_s}[u(a, \omega)].$$

The judge's default action $\hat{a}(\mu)$ is to convict if $\mu \geq (\Pi_J + \lambda)/(2\Pi_J + \lambda)$. Given this, the prosecutor's expected payoff is

$$\hat{v}(\mu) := \mathbb{E}_\mu[v(\hat{a}(\mu))] = \Pr(C|\mu)\Pi_P = \begin{cases} 0 & \text{if } \mu < \frac{\Pi_J+\lambda}{2\Pi_J+\lambda} \\ \Pi_P & \text{if } \mu \geq \frac{\Pi_J+\lambda}{2\Pi_J+\lambda}. \end{cases}$$

The value of an optimal signal is

$$\max_\tau\ \mathbb{E}_\tau[\hat{v}(\mu)] \text{ subject to } \sum_{Supp(\tau)} \mu_s \tau(\mu_s) = \mu_0,$$

where $\tau \in \Delta(\Delta(\Omega))$ is a distribution over posteriors and $Supp(\tau) = \{\mu_s\}_{s \in \{i,g\}}$. An investigation is $\pi(s|\omega) = \mu_s(\omega)\tau(\mu_s)/\mu_0(\omega)$. Finally, the concave closure of the prosecutor's expected payoff is $V(\mu) := \sup\{z : (\mu, z) \in co(\hat{v})\}$, where $co(\hat{v})$ is the convex hull of the graph of $\hat{v}$ and z is the value of a signal.

3 When Does the Prosecutor Benefit from Persuasion?

My main result is that, no matter the extent of the judge's loss aversion, the prosecutor always benefits from persuasion relative to a fully informative investigation[2]. Moreover, compared to no communication, the prosecutor benefits from persuasion whenever the judge's default action is to acquit. This result is counter-intuitive, as one would expect that beyond a threshold value of loss aversion the prosecutor would not gain from persuasion because the judge's loss from wrongful conviction would loom sufficiently large.

Theorem 1. *The prosecutor benefits from persuasion irrespective of the judge's loss aversion.*

When the judge is loss-averse, the probability of conviction under persuasion is greater than under a fully informative investigation. Therefore, as a result of persuasion, the prosecutor's expected payoff increases relative to when he leaves no uncertainty about the defendant's guilt. The prosecutor's gain from persuasion is strict when the judge is finitely loss-averse. In the limit, as loss aversion approaches infinity, the judge's conviction rate approaches the prior and, hence, the prosecutor's expected payoff approaches that obtained with a fully informative investigation. Intuitively, the benefit from persuasion is decreasing in the judge's loss aversion. The salient implication is that, as the severity of loss aversion increases, persuasion becomes less effective.

[2] A fully informative investigation leaves no uncertainty about the state and, hence, the judge convicts in line with the prior.

The judge's belief that the defendant is guilty, having received a recommendation of guilt, is $\mu_g(G) = (\Pi_J + \lambda)/(2\Pi_J + \lambda)$. This posterior belief is strictly greater than when the judge is not loss-averse[3]. The intuition is that, due to loss aversion, the judge needs to be more confident in the defendant's guilt before she is willing to convict. The value of the optimal signal to the prosecutor is

$$V(\mu_0) = \Pi_P \left(\frac{2\Pi_J + \lambda}{\Pi_J + \lambda} \right) \mu_0,$$

which is decreasing in loss aversion. The unique optimal investigation is

$$\pi(g|I) = \frac{\Pi_J}{\Pi_J + \lambda} \frac{\mu_0}{1 - \mu_0} \quad \text{with} \quad \pi(g|G) = 1.$$

Intuitively, the probability of sending a recommendation of guilt when the defendant is innocent is decreasing in the judge's loss aversion and approaches zero as loss aversion approaches infinity.

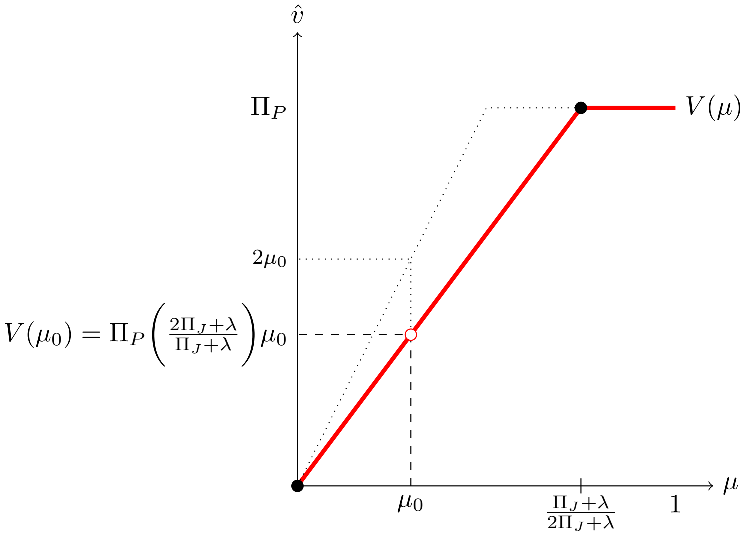

Figure 1: Persuading a loss-averse judge.

My analysis thus far has implicitly assumed that the judge's loss aversion is common knowledge. Crucially, if the prosecutor is unaware of the extent of the judge's loss aversion, persuasion may no longer increase the rate of conviction.

Corollary 1. *Persuasion is impossible if the prosecutor underestimates the judge's loss aversion.*

The intuition is that, when the prosecutor underestimates the judge's loss aversion, a false signal of guilt is sent excessively. The judge is, therefore, dubious about the signal's credibility and chooses to acquit. Conversely, when the prosecutor overestimates the judge's loss aversion, persuasion is possible; however, the prosecutor's investigation is not optimal. A greater

[3] To see this note that in Kamenica and Gentzkow (2011) this posterior is $\mu_g^{KG}(G) = 1/2$. Therefore, it is larger under loss aversion if $(\Pi_J + \lambda)/(2\Pi_J + \lambda) > 1/2$, which holds for any $\lambda > 0$.

probability of guilty recommendations would increase both the conviction rate and the prosecutor's expected payoff. Common knowledge of the judge's loss aversion is, therefore, a necessary condition for optimal persuasion.

As the prosecutor may be unable to correctly infer the judge's loss aversion without repeated interaction, this is a strong requirement. Unlike other strands of literature that study the implications of loss aversion, the presence of the kink in the judge's payoff is not sufficient. The prosecutor must also be aware of the severity of the judge's loss aversion.

One conclusion is that loss aversion may be sufficient to prevent wrongful convictions in many cases. Privately providing judges with further information on the pitfalls of wrongful conviction to strengthen loss aversion could, therefore, act as a preventative measure.

4 Quantifying the Effectiveness of Persuasion

By utilising the insights of the experimental literature that estimates loss aversion, I can compare the extent of persuasion under loss aversion to the canonical model without ad-hoc assumptions. Abdellaoui et al. (2007) estimate loss aversion and summarise a number of researchers' findings. By taking the mean of these results, I obtain an empirically supported value of $\lambda = 2.75$.

I assume $\Pi_J = \Pi_P = 1$ and $\mu_0 = 3/10$ to facilitate comparison with Kamenica and Gentzkow (2011). It follows that $\mu_g(G) = 15/19$ and $V(\mu) = (19/15)\mu$ for $\mu \leq 15/19$ or $V(\mu) = 1$ for $\mu > 15/19$. The value of the optimal signal is $V(\mu_0) = 19/50$. Therefore, compared to the canonical model, the judge's conviction rate under persuasion has fallen from sixty to thirty-eight percent[4]. This conviction rate is now only eight percent above the prior and is obtained by the prosecutor's optimal investigation

$$\begin{aligned} \pi(i|I) &= \frac{31}{35} \qquad \pi(i|G) = 0 \\ \pi(g|I) &= \frac{4}{35} \qquad \pi(g|G) = 1. \end{aligned}$$

The probability the prosecutor sends a signal of guilt when the defendant is innocent is eleven percent. This is a significant reduction from the standard model where this signal is sent with a probability of forty-two percent.

Moreover, compared to Gentzkow and Kamenica (2014), the prosecutor sends a false signal of guilt when the defendant is innocent less often under loss aversion than under costly signals, where this signal is sent with thirty-one percent probability. Yet, the conviction rate is greater under costly signals at forty-three percent.

[4] The comparison is relatively robust to changes in loss aversion. When $\lambda = 2$, the conviction rate increases to forty percent and the probability of sending a false signal of guilt rises to fourteen percent. When $\lambda = 2.5$, the probabilities are thirty-nine and twelve percent, respectively.

5 Discussion

My analysis highlights that a comprehensive study of static persuasion with receiver loss aversion would be useful. Such research would indicate the generality of my results and could yield new insights. This could include an analysis of mechanisms that may increase the judge's loss aversion and thereby decrease wrongful conviction.

Appendices

A Proofs

Proof of Theorem 1. I use results from Kamenica and Gentzkow (2011) to prove this theorem. The first is that, when the defendant is guilty, the prosecutor never sends a signal of innocence.

Proposition 1 (Kamenica and Gentzkow Proposition 4)**.** *If an optimal signal induces a belief μ that leads to a worst action, Receiver is certain of her action at μ.*

This implies that

$$\pi(i|G) = 0 \quad \text{and} \quad \pi(g|G) = 1,$$

which in turn pins down the posterior beliefs under signal i

$$\mu_i(G) = \frac{\Pr(\omega = G)\pi(i|G)}{\Pr(\omega = G)\pi(i|G) + \Pr(\omega = I)\pi(i|I)} = 0 \text{ and } \mu_i(I) = 1.$$

Therefore, under posterior μ_i, the judge's expected payoffs from each action are

$$\begin{aligned}\mathbb{E}_{\mu_i}[u(A,\omega)] &= \mu_i(G)u(A,G) + \mu_i(I)u(A,I),\\ &= \mu_i(G)u(A,G) + [1 - \mu_i(G)]u(A,I),\\ &= \Pi_J\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\mu_i}[u(C,\omega)] &= \mu_i(G)u(C,G) + \mu_i(I)u(C,I),\\ &= \mu_i(G)u(C,G) + [1 - \mu_i(G)]u(C,I),\\ &= -\lambda.\end{aligned}$$

Hence, as

$$\mathbb{E}_{\mu_i}[u(A,\omega)] > \mathbb{E}_{\mu_i}[u(C,\omega)],$$

acquitting the defendant maximises the judge's expected payoff under posterior μ_i.

The second result I use implies that the posterior μ_g is such that the judge is indifferent between convicting and acquitting.

Proposition 2 (Kamenica and Gentzkow Proposition 5)**.** *If a belief μ induced by an optimal signal is either (i) interior or (ii) leads to a best-attainable action, then Receiver's preference is not discrete at μ.*

Under posterior μ_g, the judge's expected payoffs from each action are

$$\begin{aligned}\mathbb{E}_{\mu_g}[u(A,\omega)] &= \mu_g(G)u(A,G) + \mu_g(I)u(A,I),\\ &= \mu_g(G)u(A,G) + [1-\mu_g(G)]u(A,I),\\ &= \Pi_J[1-\mu_g(G)] \end{aligned} \tag{1}$$

and

$$\begin{aligned}\mathbb{E}_{\mu_g}[u(C,\omega)] &= \mu_g(G)u(C,G) + \mu_g(I)u(C,I),\\ &= \mu_g(G)u(C,G) + [1-\mu_g(G)]u(C,I),\\ &= \Pi_J\mu_g(G) - \lambda + \lambda\mu_g(G). \end{aligned} \tag{2}$$

Equating (1) and (2) yields

$$\mathbb{E}_{\mu_g}[u(A,\omega)] = \mathbb{E}_{\mu_g}[u(C,\omega)] \ \Rightarrow\ \mu_g(G) = \frac{\Pi_J+\lambda}{2\Pi_J+\lambda},$$

which implies $\mu_g(I) = \frac{\Pi_J}{2\Pi_J+\lambda}$. Therefore, the concave closure of the prosecutor's expected payoff and the value of the optimal signal are

$$V(\mu) = \begin{cases} \Pi_P(\frac{2\Pi_J+\lambda}{\Pi_J+\lambda})\mu & \text{if } \mu \leq \frac{\Pi_J+\lambda}{2\Pi_J+\lambda} \\ \Pi_P & \text{if } \mu > \frac{\Pi_J+\lambda}{2\Pi_J+\lambda} \end{cases}$$

and

$$V(\mu_0) = \Pi_P\left(\frac{2\Pi_J+\lambda}{\Pi_J+\lambda}\right)\mu_0.$$

The unique optimal investigation is

$$\pi(g|I) = \frac{\mu_g(I)\tau(\mu_g)}{\mu_0(I)} = \frac{(\frac{\Pi_J}{2\Pi_J+\lambda})(\mu_0\frac{2\Pi_J+\lambda}{\Pi_J+\lambda})}{1-\mu_0} = \frac{\Pi_J}{\Pi_J+\lambda}\frac{\mu_0}{1-\mu_0}.$$

For finite loss aversion, the prosecutor strictly benefits from persuasion relative to a fully informative investigation as

$$\Pi_P\left(\frac{2\Pi_J+\lambda}{\Pi_J+\lambda}\right)\mu_0 > \Pi_P\mu_0 \ \Rightarrow\ \frac{2\Pi_J+\lambda}{\Pi_J+\lambda} > 1 \ \Leftrightarrow\ \Pi_J > 0$$

and relative to no communication when $\mu < (\Pi_J+\lambda)/(2\Pi_J+\lambda)$ as

$$\Pi_P\left(\frac{2\Pi_J+\lambda}{\Pi_J+\lambda}\right)\mu_0 > 0.$$

In the limit, the judge's conviction rate approaches the prior and the prosecutor's expected payoff approaches that under a fully informative investigation as

$$\lim_{\lambda\to\infty} \Pi_P \left(\frac{2\Pi_J + \lambda}{\Pi_J + \lambda} \right) \mu_0 = \lim_{\lambda\to\infty} \Pi_P \left(\frac{\frac{2\Pi_J}{\lambda} + 1}{\frac{\Pi_J}{\lambda} + 1} \right) \mu_0 = \Pi_P \mu_0.$$

□

Proof of Corollary 1. Suppose the prosecutor misperceives the judge's loss aversion as $\tilde{\lambda} := \beta\lambda$ for $1 > \beta \geq 0$. The posterior $\mu_g^\beta(G)$ is found by

$$\Pi_J[1 - \mu_g^\beta(G)] = \Pi_J \mu_G^\beta(G) - \beta\lambda + \beta\lambda\mu_G^\beta(G) \;\Rightarrow\; \mu_g^\beta(G) = \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda}.$$

Substituting $\mu_g^B(G)$ into (1) and (2) implies the judge's optimal action is to acquit if

$$\mathbb{E}_{\mu_g^\beta}[u(A,\omega)] = \Pi_J - \Pi_J \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda} > \Pi_J \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda} - \lambda + \lambda \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda} = \mathbb{E}_{\mu_g^\beta}[u(C,\omega)],$$

which is equivalent to

$$\Pi_J + \lambda > 2\Pi_J \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda} + \lambda \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda},$$

$$\Pi_J + \lambda > (2\Pi_J + \lambda) \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda},$$

$$\frac{\Pi_J + \lambda}{2\Pi_J + \lambda} > \frac{\Pi_J + \beta\lambda}{2\Pi_J + \beta\lambda},$$

$$0 > (\beta - 1) \frac{\lambda\Pi_J}{(2\Pi_J + \beta\lambda)(2\Pi_J + \lambda)}. \tag{3}$$

Condition (3) holds by the assumption that $1 > \beta \geq 0$. □

References

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. Loss Aversion under Prospect Theory: A Parameter-Free Measurement. *Management Science*, 53:1659–1674, 2007.

Garoupa, N. and Rizzolli, M. Wrongful Convictions do Harm Deterrence. *Journal of Institutional and Theoretical Economics*, 168:224–231, 2012.

Gentzkow, M. and Kamenica, E. Costly Persuasion. *American Economic Review: Papers and Proceedings*, 104:457–462, 2014.

Kamenica, E. and Gentzkow, M. Bayesian Persuasion. *American Economic Review*, 101:2590–2615, 2011.

Lipnowski, E. and Mathevet, L. Disclosure to a Psychological Audience. *American Economic Journal: Microeconomics*, 10:67–93, 2018.

Liu, X. Disclosing Information to a Loss-Averse Audience. *Economic Theory Bulletin*, 6:63–79, 2018.

Mariotti, T., Schweizer, N., Szech, N., and von Wangenheim, J. Information Nudges and Self-Control. 2018.

Matysková, L. Bayesian Persuasion with Costly Information Acquisition. 2018.

Silbert, R., Hollway, J., and Larizadeh, D. Criminal Injustice: A Cost Analysis of Wrongful Convictions, Errors, and Failed Prosecutions in California's Criminal Justice System. Technical report, University of California, Berkeley, 2015.

Traynor, R. J. *The Riddle of Harmless Error*. Ohio State University Press, 1970.