# Private Information in Two-Dimensions: The Case of Malicious Plaintiffs

Brishti Guha
Jawaharlal Nehru University

Shay Lavie *
Tel-Aviv University

August 15, 2025

**Abstract**

The paper explores two-dimensional private information in pre-trial bargaining. Traditional models of litigation assume private information in a single dimension, where that information affects the expected judgment at trial (e.g., damages). But in reality, informed parties could also diverge on other dimensions, and in particular, concerning properties that do not affect the expected judgment at trial.

While this phenomenon has various manifestations in real-world setups, we utilize the example of malicious motivations. Plaintiffs could diverge on their damages, but they can also have different motivations to litigate. While the different motivations affect the plaintiffs' "appetite" for trial, they do not affect the expected judgment at trial. Our analysis reveals that the presence of this second dimension, informational gaps regarding the plaintiff's motivations, leads to surprising effects. First, the familiar fully-revealing result is not sustainable with bi-dimensional private information. Those who have weaker "appetite" for trial tend to "reverse mimic" to the weakest type (and settle with certainty). This effect may increase the rate of settlements. Second, the uninformed party is counterintuitively better off — intuitively, the addition of private information in the second dimension allows her to be more aggressive, which facilitates more generous offers. Third, bi-dimensional private information gives rise to various equilibria, semi-fully-revealing, binary (accept/reject), and semi-pooling ones.

Our results bear empirical predictions. We predict "focal" settlement points, especially at the low-end of the distribution of types; challenge the assumption regarding monotonicity in settlement rates; and offer a rational explanation for over-rejection and overly generous offers observed in experiments.

**Keywords:** pre-trial bargaining, litigation, malicious motivations, signaling.
**JEL classification:** K41, K40

*We thank ***.

# 1. Introduction

It is well-known that the vast majority of the cases settle (e.g., Spier (2007), p. 268). However, the cases that do not settle pose significant costs, to the parties and the judiciary as well (e.g., id., at 262-64; Shavell (2004), p. 281). Indeed, "[j]udiciaries worldwide are grappling with increasing caseloads" (Engel and Weinshall (2020), p. 722). The skyrocketing expenses of legal proceedings have led to voluminous literature that explored pre-trial bargaining and settlement failures. In particular, asymmetric information — namely, where one of the parties possesses private information — has emerged as a significant source of settlement failures.

We seek to contribute to this literature by introducing the case of two-dimensional private information. Specifically, we consider two sets of private information. The first relates to the personal attributes of the informed parties, which affect the expected judgment at trial. Straightforward examples are information concerning damages and liability. The literature has extensively discussed private information along this dimension. The second dimension, however, relates to the idiosyncratic attributes of the parties that affect their payoff from trial, but have no impact on the result at trial. Perhaps surprisingly, we have not come across previous literature that analyzes private information in this second dimension. We believe that private information in this second dimension is a common feature in daily pre-trial bargaining situations. As we will show below, introducing private information in this second dimension leads to counterintuitive results concerning the prediction of traditional models.

To illustrate, we will use throughout the paper the example of (possibly) malicious plaintiffs. It is common to depict plaintiffs as having private information with regard to their expected damages: expected damages differ among plaintiffs, but the defendant knows only the distribution of the damages in the population of the plaintiffs rather than the expected damages of her specific rival (e.g., the seminal Reinganum and Wilde (1986)). These are the familiar informational gaps in the first dimension, that do affect the expected judgment. But plaintiffs could also be motivated by malicious reasons. In that case, they gain from their rival's loss, although their malicious motivations do not affect the expected judgment at trial (e.g., Guha (2016); for empirical findings that suggest malicious motivations see, e.g., Deck et al. (2024)). Malicious motivations demonstrate, then, our second dimension — personal properties that affect litigants' outside options but not the expected judgment at trial. More generally, the plaintiff's true motivations may, of course, be her private information. One can think, then, of private information along these two dimensions — expected damages *and* malicious motivations.

We note that the specific case of malice is by no means the only possible example of a two-dimensional setup. One can think of other examples where there is private information that relates to the expected value of the case, e.g., severity of damages; and additional private information that affects the payoff of the informed party without affecting the expected value of the case directly. A straightforward example is the trial expenses of the plaintiff (should the case go to trial). The literature has assumed that plaintiffs have similar trial expenses; but they could well differ with respect to their costs of trial. Some plaintiffs could well have easier access to lawyers, and/or experience and skill that enable them to better monitor their lawyer and cut legal expenses. These differences could be the plaintiffs' private information, where plaintiffs with lower trial costs have a stronger "appetite" for trial (but arguably similar expected judgment at trial), similar to our leading example of malicious motivations.

We explore the two-dimensional setup through a model in which plaintiffs are informed with respect to both their damages (along a continuum) *and* their malicious motivations (non-malicious and malicious plaintiffs). The plaintiff proposes a single take-it-or-leave-it (TIOLI) offer to the defendant, and, should the latter reject, the case goes to trial. If all plaintiffs are non-malicious, we predict a fully revealing equilibrium, where the defendant utilizes an acceptance function which decreases in the settlement offer to prevent mimicking by weak plaintiffs and maintain equilibrium (Reinganum and Wilde (1986)). Where all plaintiffs are malicious, the resulting equilibrium is similar in kind. However, as malicious plaintiffs have a greater "appetite" for trial (they derive more utility thereby), the defendant rejection behavior is more aggressive: intuitively, where one's appetite for trial is larger, one needs to reject more offers to prevent mimicking and maintain equilibrium.

We then integrate malicious and non-malicious plaintiffs. If the defendant faces either a non-malicious *or* a malicious plaintiff, she should utilize the more parsimonious acceptance function — otherwise, she invites mimicking by weak, malicious plaintiffs, who are not deterred by the acceptance rate that deters weak, non-malicious plaintiffs. The tendency of the defendant reject more aggressively (given the possibility that these offers come from malicious plaintiffs) leads some non-malicious plaintiff to "reverse-mimic," that is, pretend as weaker types and propose a more generous offers. Reverse-mimicking allows the non-malicious to settle more frequently.

Our main takeways could be summarized as follows. First, perhaps counterintuitively, the introduction of additional informational gaps, in this second dimension, could lead to more settlements. This effect stems from the foregoing, cautious acceptance behavior of the uninformed, given the existence of two-dimensional informational gaps. This acceptance behavior results in "reverse-mimicking" to weaker types, and accordingly, more settlements. Thus, in addition to higher settlement rates, the introduction of a second dimension of private information could surprisingly benefit the uninformed party too. This benefit comes at the expense of the privately informed type who has a lesser "appetite" for trial, in our example, non-malicious plaintiffs. Intuitively, the existence of (unobservable) malicious types allows the uninformed to commit to be more aggressive towards non-malicious types in rejecting settlement proposals. This aggressive acceptance rate pushes the non-malicious to propose more generous offers. However, as we explain in greater detail in the paper, the overall settlement rate may or may not be correlated with malice.

Second, and relatedly, our results undermine the familiar fully-revealing equilibrium that the literature predicts (starting with the seminal Reinganum and Wilde (1986)). This follows from the unambiguous tendency of non-malicious types to reverse-mimic to the weakest type (whose offer is accepted with certainty). Instead, we show three possible equilibria. In the first, all non-malicious types reverse-mimic to the weakest type and always settle; whereas malicious types reveal, similarly to the standard predictions. We thus denote this equilibrium as "semi-fully-revealing." However, this state is not sustainable where the continuum of plaintiff types stretches further. In that case, weaker plaintiffs (non-malicious and malicious alike) reverse-mimic to the weakest type (and always settle) whereas the stronger always litigate. As the defendant either accepts or rejects with certainty, we refer to this state as a "binary" equilibrium. The third equilibrium extends from the binary state. Weaker non-malicious types reverse-mimic to the weakest types (as in previous cases); intermediate malicious and non-malicious plaintiffs pool on an intermediate offer; and stronger types always litigate (as in the binary equilibrium). As a mass of intermediate plaintiffs pools and proposes an intermediate offer, we refer to this equilibrium as "semi-pooling."

Our analysis generates clear empirical predictions as well as different interpretations of current empirical findings. First, we predict "focal" points of settlements — unambiguously, at the low end (weakest settlement), and possibly another "focal," intermediate settlement (on which the parties pool in the semi-pooling state). Second, we predict that settlements are over rejected relative to the standard predictions. The reason is the inability of the uninformed type to decipher the outside gain of the informed party (namely, malice), as well as its desire to prevent mimicking by those with greater appetite for trial. Expecting this aggressive behavior, the informed party which has a smaller appetite for trial proposes more generous offers. Interestingly, these predictions are consistent with experimental findings that suggest that participants, on the one hand, over-reject (relative to standard predictions) and, on the other hand, propose overly generous offers (cf., Pecorino and Van Boening (2018), Guerra et al. (2025)). Another important implication relates to the assumption of monotonicity in acceptance rates, that is, that stronger cases are more likely to litigate under asymmetric information. The assumption of monotonicity enables us to infer from litigated cases, selection effects notwithstanding (Klerman and Lee (2014)). However, we show that the introduction of an additional dimension of private information could flip this prediction (provided that those with smaller appetite for trial have better cases).

**Related Literature.** Our paper builds of course on the classic literature on pre-trial bargaining under asymmetric information, that is, where one party has private information (Bebchuk (1984), Reinganum and Wilde (1986)). This literature has spawned various branches of research, including work on two-sided private information, e.g., where the plaintiff (defendant) is privately informed with regard to damages (liability) (e.g., Schweizer (1989), Daughety and Reinganum (1994), Dari-Mattiacci and Saraceno (2020), Abramowicz (2024); for example in the criminal context see Reinganum (1988)). Unlike the two-sided asymmetric information literature, our paper assumes that only one of the parties has private information, but adds another dimension of private information, assuming it affects only the outside option of the informed party.

Our results pertain to the predictions of the familiar asymmetric information models regarding settlements rates and patterns (e.g., discontinutiy of settlements and mass of settlements at the lower end). Hence, our paper relates to empirical and experimental literature on pre-trial bargaining (for a brief survey see Klerman and Lee (2014), pp. 209–211; for experimental literature see Pecorino and Van Boening (2018)). Moreover, our results indicate that the assumption of monotonicity, compare Klerman and Lee (2014), may not be true where we consider the second dimension, namely, appetite for trial.

Finally, our analysis utilizes the specific case of malicious motivations to analyze private information in two dimensions. Several works have explored the role of malice in pre-trial bargaining. Similarly to our paper, Guha (2016) assumes malicious motivations that raise litigants' payoffs from trial without directly affecting its results. Under complete information, Guha shows that malicious motivations inhibit settlements and affect their timing (id; Guha (2019)). Importantly for our purposes, recent empirical literature has shown that the incorporation of malicious motivations helps explain experimental findings, e.g, that plaintiffs with negative expected-value cases go to trial even where their offers are refused (Deck et al. (2024)). This literature highlights the importance of discussing two-dimensional private information, as we do in our paper, and its potential empirical importance.

**Plan of the Paper** We start in Section 2 with a simple numerical example in a two-type case. We use this simple example to motivate the discussion and illustrate the reverse-mimicking phenomenon. We then move to our main model. Section 3 presents our setup. We build on the seminal continuous model in which the informed plaintiff proposes a take-it-or-leave-it offer to the uninformed defendant (cf. Reinganum and Wilde (1986)). To this model we introduce private information in the second dimension, namely, malicious motivations. Section 4 lays out the benchmark case, that is, where the plaintiff's malice is common knowledge. We show that in equilibrium the acceptance function that the uninformed party utilizes against malicious plaintiffs is more aggressive than the acceptance function in the non-malicious case. Intuitively, where plaintiffs have greater appetite for trial (that is, they are driven by malice), there should be more rejection to prevent mimicking and maintain equilibrium. Section 5 integrates non-malicious and malicious plaintiffs, assuming that malice is the plaintiff's private information. We show that there are three possible resulting equilibria: semi-fully-revealing, binary, and semi-pooling. Before concluding, in Section 6 we summarize and discuss our results. We also briefly discuss the same dynamics under different setups, e.g., where the informed party does not enjoy the entire settlement surplus, and where there are alternative forms of private information in the second dimension (that is, affecting the outside option but not the expected judgment at trial). The Appendix contains some of the relevant proofs.

## 2. Exposition: Two Type Example

Before discussing our main model we start with a simple numerical example in a two-type setup. Following previous asymmetric information models, suppose that there are two possible plaintiffs, with low and high expected damages, respectively 30 and 55. The damages are the plaintiff's private information, where the defendant only knows that these two types of plaintiffs are equally represented in the population. The plaintiff makes a single settlement offer to the defendant, and if the offer is rejected the case goes to trial where the court finds out the actual damages. If the case goes to trial, the plaintiff pays 30 and the defendant 20 as trial expenses. In addition to this standard description now suppose that there are two types of plaintiffs, again, equally represented in the population: malicious and non-malicious, where each can have either low or high damages. Malicious types incur additional utility, denoted $\mu$, from each dollar that the defendant pays, through settlement, trial, or trial expenses. We will first present the standard analysis for malicious and non-malicious types, and then integrate them.

*Non-Malicious Plaintiffs.* Suppose the plaintiff knows that she faces a non-malicious plaintiff. The standard prediction is full-revelation. Low-damages plaintiffs offer 50. As this sum equals the defendant's liability at trial, inclusive of trial costs, these offers are taken by the defendant. High-damages plaintiffs offer 75, which expresses the defendant's expected liability from going to trial against them. But to maintain equilibrium, and prevent low-damages plaintiffs from masquerading as high-damages ones, the defendant has to take some high offers to trial. One can verify that by taking $p = 2/3$ of the high offers to trial the low-damages types are just indifferent between revealing their type and settling for 50; and proposing 75 and risking a trial in which their net payoff is zero.

*Malicious Plaintiffs.* Suppose the defendant knows that she faces a non-malicious plaintiff. Malicious plaintiffs have a stronger appetite for trial. If a low-damages malicious plaintiff goes to trial, she expects $30(1 + \mu) + 20\mu - 30 = 50\mu$ (compared to zero for the non-malicious low-damages plaintiff). However, the plaintiff's malice has no direct effect on the expected value at trial, that is, on the defendant's liability. We should therefore predict a similar fully-revealing equilibrium with the same settlement offers. The low-damages malicious plaintiff proposes 50 and settles with certainty. High-damages malicious plaintiffs propose 75 and are taken to trial sufficiently often to prevent mimicking by low-damages malicious plaintiffs. The rate at which offers of 75 are taken to trial, denoted $p_M$ could again be calculated by the low-damages malicious plaintiff indifference function. If she reveals and settles, she gains $50(1+\mu)$. If she mimics and offers 75, she expects $75(1+\mu)p_M + (1-p_M)50\mu$. One can verify that $p_M = \frac{2}{3+\mu}$, which is necessarily lower than the previous acceptance rate, 2/3. Intuitively, as malicious types have a greater appetite for trial, one needs a lower acceptance rate to maintain equilibrium.

*Two Dimensions.* What happens where the defendant lacks information both on the expected damages *and* the plaintiff's malice, representing here appetite for trial? Both malicious and non-malicious plaintiffs with low damages, as we have shown, offer 50. The plaintiff's malice does not affect the defendant's expected liability, hence these offers are still accepted with certainty. However, the analysis is more nuanced where the defendant faces a high offer, of 75. With full information on spite, in equilibrium the defendant would have accepted offers of 75 from malicious types with $p_M = \frac{2}{3+\mu}$ and from non-malicious types with 2/3. But without information on malice, the defendant is worse off accepting offers of 75 at a rate lower than $p_M = \frac{2}{3+\mu}$. By doing so, the defendant invites malicious, low-damages plaintiff to offer 75, as the latter are no longer indifferent between mimicking and revealing. Hence the defendant should respond to offer of 75 with $p_M = \frac{2}{3+\mu}$. However, this acceptance rate worsens the situation of high-damages non-malicious plaintiffs — as they are taken to trial more often than in the benchmark case, due to the inability of the defendant to decipher malice.

Suppose that $\mu = 0.5$. Under the benchmark case, the high damages non-malicious plaintiffs expect, given the $\frac{2}{3}$ acceptance rate, $2/3 * 75 + 1/3(55 - 30) = 58\frac{1}{3}$. With the additional asymmetric information concerning spite, $p_M$ decreases from $\frac{2}{3}$ to $\frac{4}{7}$, and the expected gain for this type of plaintiff likewise decreases to $53\frac{4}{7}$.

Suppose now a higher malice factor $\mu = 1$. Accordingly, $p_M$ further decreases to 0.5, and the expected gain of the non-malicious high-damages type from trial is only 50. Now, due to the higher malice factor, the high-damages non-malicious plaintiff is (weakly) better off offering 50, like the low-damages plaintiffs, and settling with certainty. We refer to this dynamics as reverse-mimicking.

Moreover, where the high-damages non-malicious type reverse-mimics and pools with the (malicious and non-malicious) low-damages plaintiffs, these three types can raise their offer beyond 50. Recall that we assumed equal proportions of low- and high-damages plaintiffs, as well as malicious and non-malicious ones. Thus, where the pooled low offer is $58\frac{1}{3}$, the defendant is just indifferent between accepting it or going to trial and facing a 55 type with probability 1/3 and a 30 type with probability 2/3.

The appendix provides a full characterization and proof of the two type case. Where the malice factor is sufficiently high with respect to the damages, high-damages non-malicious types reverse mimic, and pool together with low types (both malicious and non-malicious). If the malice factor is not sufficiently high, the non-malicious type is bound to offer her original, revealing offer; but goes to trial more often than in the benchmark case. In this sense, the defendant is better off at the expense of the non-malicious high-damages plaintiff, due to the former's inability to decipher malice. While the core dynamics remains similar in kind, below we provide a full characterization and more formal discussion, using the setup of a continuum of plaintiff types.

## 3. Setup: Continuous Case

We now turn to explore more carefully the foregoing dynamics. We assume that the plaintiff is privately informed regarding two dimensions: her expected damages $J$ and her malicious motivations $M \in \{0, 1\}$. The sequence of the game is identical to classic pre-trial bargaining games. At the first stage the plaintiff proposes a single take-it-or-leave-it offer (TIOLI) $S$. The defendant responds by either accepting or rejecting ($R : (S) \rightarrow \{Accept, Reject\}$). If the offer is accepted, the payoffs are determined by the settlement offer. In case the offer $S$ is rejected and the case goes to trial, the trial expenses of each party are $c_D$ for the defendant and $c_P$ for the plaintiff. We denote the sum of trial costs as $T = c_P + c_D$.

While the expected damages of each plaintiff are her private information, these damages are distributed along a continuum in the population of each group of plaintiffs (malicious and non-malicious), and that distribution is common knowledge. We assume that malicious and non-malicious plaintiffs share the very same distribution of expected damages. For tractability, we let the damages of the weakest plaintiff (malicious and non-malicious alike) be equal to $c_P$, her trial expenses; whereas the damages of the strongest plaintiff are equal to $\bar{J}$. The cumulative distribution function of the damages is $F$.

In addition to expected damages, the plaintiff is also privately-informed concerning another dimension, her inner motivation: malicious or non-malicious (NM). The proportion of malicious plaintiffs, denoted $\beta$, is common knowledge. Importantly, malicious motivations affect the monetary payoff of the plaintiff: malicious types enjoy a (commonly known and positive) fraction $\mu$ of the rival's loss (from judgment, trial expenses, or settlement). A higher $\mu$ thus reflects a more significant spite from the plaintiff's side, and a larger gain from the rival's loss. Thus, if the defendant settles for $S$, the malicious plaintiff profits $S(1+\mu)$; if the defendant goes to trial and pays $J+c_D$, the malicious plaintiffs enjoys $J(1+\mu)+c_D\mu$ (minus her own trial expenses). We assume that these two dimensions of private information, damages and malicious motivations, are uncorrelated with each other.

By construction, there are no plaintiffs in our setup with negative expected-value. As the damages of the weakest type are $c_P$, the weakest NM plaintiff has a precise profit of zero from trial; whereas the weakest malicious type gains $c_P(1+\mu)-c_P+\mu c_D = T\mu$. We also assume that all parties are risk-neutral. We refer to the beliefs of the defendant through the function $(B : S \longrightarrow J_i)$, which assigns a type $J_i$ to each offer. We use Perfect Bayesian Nash Equilibrium as our solution concept.

## 4. Benchmark Case: One Dimension

We will first briefly discuss the simple case in which there is a single dimension of private information, concerning the expected damages at trial. We will start with the case of NM plaintiffs only (with private information concerning their damages) and proceed to the case of only malicious plaintiffs (again, with private information regarding their damages).

### 4.1. Non-Malicious Plaintiffs

The well-known solution to the case of privately-informed, NM plaintiffs results in a fully-revealing equilibrium (Reinganum and Wilde (1986). Each plaintiffs proposes a TIOLI offer that fits the costs of the defendant from taking her to trial, namely, $S_i = J_i + c_D$. This equilibrium is maintained by the defendant's acceptance function $p(S)$. Intuitively, to prevent mimicking to stronger types the defendant should take higher offers to trial more often. To find the acceptance function $p(S)$ that maintains complete revelation consider first the payoff function of the NM plaintiff, $\pi_{NM}$:

$$\pi_{NM}(S) = p(S)S + (1-p(S))(J-c_P).$$

We can differentiate this payoff function with respect to $S$, and equate the derivative to zero when evaluated at the truthful offer, $S^* = J + c_D$. This results in the following first-order conditions:

$$p(S) = -p'(S)T.$$

The general solution to the above differential equation is $p(S) = Ce^{\frac{-S}{T}}$. We can find the constant $C$ by solving for the boundary condition, namely, as the lowest offers (from the weakest types) are accepted with $p(S) = 1$. This yields the following acceptance function:

$$p(S) = e^{1-\frac{S}{T}}. \quad (1)$$

Observe that $S \geq T$ and thus $p(S)$ declines with $S$. Intuitively, to maintain equilibrium, higher offers should be taken to trial more often.

### 4.2. Malicious Plaintiffs

We can now sketch the parallel case, where the plaintiffs are only privately informed with respect to their damages, but, unlike the previous setup, all plaintiffs have malicious motivations. As before, we can find a fully-revealing equilibrium, in which the defendant's acceptance function is now denoted $p_M(S)$. Observe that, as before, $S^* = J + c_D$. This stems from the fact that malice does not affect the amount that the defendant has to pay at trial. However, the malicious motivations of the plaintiff change her payoff function, as now the plaintiff gains an additional amount from the defendant's loss:

$$\pi_M(S) = p_M(S)(1+\mu)S + (1 - p_M(S))(J(1+\mu) + \mu c_D - c_P).$$

As before, we can find the acceptance function $p_M(S)$ that maintains revelation by differentiating with respect to $S$, and equating the derivative to zero when evaluated at $S^* = J + c_D$:

$$(1+\mu)p_M(S) = -p_m^{'}(S)T.$$

The general solution to this differential equation is $p(S) = Ke^{\frac{-S(1+\mu)}{T}}$. And again we can find the constant $K$ by the boundary condition and obtain the following acceptance function:

$$p(S) = e^{(1+\mu)(1-\frac{S}{T})}. \quad (2)$$

As before, observe that $1 - \frac{S}{T}$ is negative. Therefore, $p_M$ again declines with $S$. Moreover, observe that where $\mu$ is higher, that is, plaintiffs are more spiteful, the rate of rejection increases. Intuitively, as spite increases, trial looms more attractive, and the incentive to mimic to a higher type (and risk a trial) is larger. Then, a higher rate of rejection is required to prevent mimicking.

### 4.3. Numerical Example

We have shown that, where the malice factor is known, malice leads to a lower acceptance rate. The following figure, Figure 1, shows that indeed the defendant rejects more offers where they come from malicious types, namely, $p_M(S) < p(S)$. We assume that trial expenses for the plaintiff and defendant are 30 and 20, respectively, and that the malice factor $\mu$ equals 0.5. The x-axis represents the offer, from 50-100 (hence, the weakest and strongest plaintiffs have expected judgments of 30 and 80, respectively). The y-axis shows the rate of acceptance. Malicious (NM) types are represented with blue (orange) curve. As can be seen from Figure 1, the rate of acceptance naturally declines with the settlement offer. More interestingly, offers who come from (known) malicious types are less likely to be accepted than the parallel offers of non-malicious types:

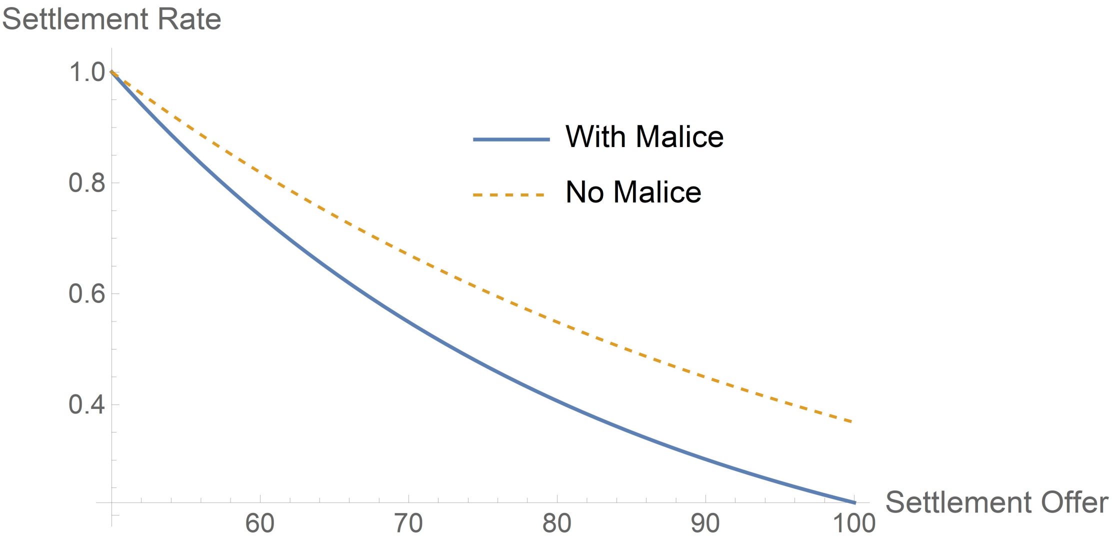


Figure 1: Acceptance Rates: With and Without Malice

## 5. Two-Dimensions

We now integrate these two dimensions of private information – damages and malice. While these dimensions are the plaintiff's private information, the defendant only knows the distribution of damages and the proportion of malicious plaintiffs in the population. As we will show below, while the equilibrium dynamics for each group of plaintiffs – malicious and non-malicious – is similar in kind, their integration yields substantially different results.

First, observe that a revealing offer from a type $J_i$ equals $J_i + c_D$ regardless of its malice (as malice has no effect on the defendant's loss at trial). Accordingly, the weakest type, $\underline{J} = c_P$ offers $\underline{S} = T$ regardless of her malice. This offer is accepted with certainty, implying that this type has no incentive to deviate.

However, the differences regarding spite do affect other types, beyond the weakest one. With full information on spite, in equilibrium the defendant would have accepted offers from malicious (NM) with $p_M$ ($p$), where $p(S) > p_M(S)$). Lacking this information, the defendant is worse off responding to an offer $S$ with the any acceptance rate that is more generous than $p_M(S)$, and in particular $p(S)$. Such a generous rate of acceptance induces mimicking by malicious plaintiffs, as only $p_M(S)$ (and less generous acceptance functions) can deter mimicking by malicious types. Mimicking, in turn, unambiguously hurts the defendant – under mimicking, the defendant unnecessarily pays higher settlements more often, where she could have benefited by rejecting some of them and going to trial. The defendant thus responds to $S$ with $p_M$. However, this acceptance rate is too low for the *non-malicious* types to induce revelation. As the Exposition has shown, these type are induced to "reverse-mimic" as lower types. This dynamics leads to interesting equilibria.

Accordingly, we will discuss three possible equilibria. In the first, the tendency of the NM plaintiffs to reverse-mimic pushes all of them to propose the lowest possible offer, $T$ (which is accepted with certainty). All malicious plaintiffs, though, reveal as in the one-dimensional analysis. We thus refer to this equilibrium as "semi-fully-revealing." In the second equilibrium, the defendant cannot maintain a fully-separating equilibrium even within the malicious plaintiffs. Thus, she either accepts with certainty (low offers) or rejects (other offers). Accordingly, weak plaintiffs with or without spite, and intermediate nonmalicious plaintiffs, propose the lowest possible offer. As the defendant either accepts or rejects we refer to this equilibrium as a "binary" equilibrium. The third equilibrium directly extends from the binary state. While weak non-malicious plaintiffs reverse-mimic to the lowest offer and settle with certainty (as in the semi-fully-revealing case), intermediate plaintiffs, malicious and non-malicious, pool together and offer an intermediate settlement (which is sometimes rejected). Stronger types, if they exist, prefer trial. We will discuss these states in order.

### 5.1. Semi Fully-Revealing

We will construct an equilibrium in which, given the inability to decipher spite, the defendant can respond with $p_M(S)$. The payoff of non-malicious plaintiffs as a function of $S$ where the defendant responds with the parsimonious acceptance rate, $p_M$, will be denoted $\pi_{NM}(S, p_M)$:

$$\pi_{NM}(S, p_M) = p_M(S) * S + (1 - p_M(S))(J - c_P). \tag{3}$$

The derivative of this acceptance function with respect to $S$ is:

$$\frac{\partial \pi_{NM}}{\partial S} = \frac{p_M(S)\left[(J - S)(1 + \mu) + c_D - c_P\mu\right]}{T}. \tag{4}$$

As $p(S)$ just deters NM plaintiffs from mimicking, the higher rate, $p_M(S)$ likewise prevents mimicking by non-malicious types. Intuitively, it also shifts NM plaintiffs further away from revelation, that is, it induces them to mask as *weaker* types, what we term as "reverse-mimicking." More formally:

**Lemma 1.** *Non-malicious plaintiffs $J \in (\underline{J} = c_P, \bar{J}]$ will always prefer to deviate downwards from truthful revelation.*

*Proof.* Consider a non-malicious plaintiff $J_i$ who truthfully reveals her type and offers $S = J_i + c_D$. Given that the defendant responds with $p_M$, the payoff of that plaintiff is given by $\pi_{NM}(S, p_M)$ (Equation 3). The derivative of $\pi_{NM}(S, p_M)$ with respect to $S$ is given by Equation 4. Evaluated at the revealing offer, this derivative equals $\frac{\partial \pi_{NM}}{\partial S_{|S=J_i+c_D}} = -\mu p_M(S) < 0$, implying that the non-malicious plaintiff is always tempted to reverse mimic to a lower type. ■

Consider again the derivative of the payoff function with respect to $S$, Equation 4. Note that $p_M$ and $T$ are non-negative. Hence, whether this derivative is positive or negative (that is, whether the non-malicious plaintiff should increase or decrease $S$)

depends on the bracketed part, $(J - S)(1 + \mu) + c_D - c_P\mu$, which linearly depends on $J - S$. This implies that for sufficiently high values of $J$, Equation 4 becomes positive. Accordingly, we will denote as $\hat{J}(S)$ the cutoff type above which the non-malicious plaintiff who offers $S$ is better off increasing her offer:

$$\hat{J}(S) = S + c_P - \frac{T}{1+\mu}. \tag{5}$$

Below we establish that all NM plaintiffs up to $\hat{J} = \hat{J}(T) = c_P + T\frac{\mu}{1+\mu}$ reverse mimic to the lowest offer, $\underline{S} = c_P + c_D = T$.

**Lemma 2.** *All non-malicious plaintiffs $J_i \in (\underline{J} = c_P, \hat{J} = c_P + T\frac{\mu}{1+\mu}]$ reverse mimic to the lowest offer $T$.*

*Proof.* First, observe that the weakest NM plaintiffs, in the vicinity of $\underline{J} = c_P$, are better off reverse-mimicking to $T$. This follows from 1, and can also be seen by evaluating Equation 4 at $J = c_P$ and obtaining a negative sign.

Second, observe from 5 that the derivative of the NM plaintiff's payoff is no longer negative where $J \geq c_P + T\frac{\mu}{1+\mu} = \hat{J}$ and $S = T$. This implies that where $J \geq \hat{J}$ NM plaintiffs are better off raising their offer above the lowest one $\underline{S} = T$. ■

We can now fully characterize the resulting, semi fully-revealing equilibrium.

**Proposition 1.** *Suppose $\bar{J} \leq \hat{J}$. Then, a perfect Bayesian equilibrium exists such that:*

*(i)* $$S^*(J) = \begin{cases} J + c_D & \text{if } M = 1 \\ T & \text{if } M = 0 \end{cases}$$

*(ii)* $$R^*(S) \longrightarrow \{Accept, Reject\} \begin{cases} (p_M, 1 - p_M) & \text{for } T < S \leq \bar{S} \\ (1, 0) & \text{for } S \leq T \\ (0, 1) & \text{elsewhere.} \end{cases}$$

*(iii)* $$B^*(S) = \begin{cases} S - c_D & \text{if } T < S \leq \bar{S} \\ \int_{c_P}^{\hat{J}} (S - c_D) dF & \text{if } S = T \\ \bar{J} & \text{elsewhere.} \end{cases}$$

*Where:* $\bar{S} = \bar{J} + c_D; \quad p_M = e^{(1+\mu)(1-\frac{S}{T})}; \quad \hat{J} = c_P + T\frac{\mu}{1+\mu}$

*Proof. i)* First, consider malicious plaintiffs ($M = 1$). Since these plaintiffs' demands are accepted according to $p_M(S)$, they are induced to reveal truthfully and their payoff is maximized by doing so. Thus, they have no incentive to deviate. If they offer above $\bar{S}$, their offer is rejected for sure. If they offer below $\underline{J} = T$ (which is accepted with certainty), they can offer precisely $T$ and obtain certain acceptance.

*ii)* Second, consider non-malicious plaintiffs ($M = 0$). As by assumption $\bar{J} \leq \hat{J}$, by Lemma 2 all of them prefer reverse-mimicking to $T$.

*iii)* Finally, consider the defendant. When a defendant faces an offer $S \in (T, \bar{S}]$, it believes that it was proposed by a malicious type. Hence, the defendant happily utilizes $p_M(S)$ to accept that offer. As $p_M(S)$ induces revelation among malicious types, the defendant is indifferent between accepting and rejecting in that case. When she faces an offer $T$, the defendant is better off accepting it, as at trial she will have to pay at least that amount. ■

Perhaps counterintuitively, the introduction of asymmetric information in the second dimension actually encourages settlements — now, all non-malicious plaintiffs prefer to reverse-mimic and obtain certain acceptance. Nonetheless, observe that these plaintiffs are unambiguously worse off compared to the one-dimensional case. Indeed, in the one-dimensional case they preferred to reveal and offer $J_i + c_D$ (and risk trial with $1 - p(S)$) over certain acceptance of an offer $T$. Malicious plaintiffs, by contrast, are indifferent, as for them the resulting equilibrium is identical. The defendant is unambiguously better off, as she manages to extract some surplus from non-malicious types. The inability to tell the plaintiff's spite, in essence, enables the defendant to be committed to reject more aggressively, pushing the non-malicious types to reverse-mimic.

Finally, we stress that this equilibrium assumes that the range of plaintiffs is restricted, $\bar{J} \leq \hat{J} = c_P + T\frac{\mu}{1+\mu}$. As we will show below (Section 5.2), where this condition does not hold the resulting equilibrium is quantitatively different. The cutoff $\hat{J}$ seems restrictive. Although a higher malice factor can stretch $\hat{J}$ upwards, this cutoff is still bound by $c_P + T$, where $\mu$ approaches infinity. This suggests that the gap between the weakest and the strongest plaintiff cannot be greater than the sum of trial costs. We can demonstrate by using the parameters of Figure 1: where the plaintiff and defendant's trial expenses were 30 and 20, respectively, and the malice factor is assumed to be 0.5. In that case, the foregoing equilibrium depends on $\bar{J} < 46.67 = 30 + 50 * \frac{0.5}{1.5}$. We demonstrate the resulting equilibrium in Figure 2 below, which shows the acceptance rates of malicious (blue) and NM (dashed orange) plaintiffs in this numerical example, where the strongest plaintiff has an expected value of 46.67:

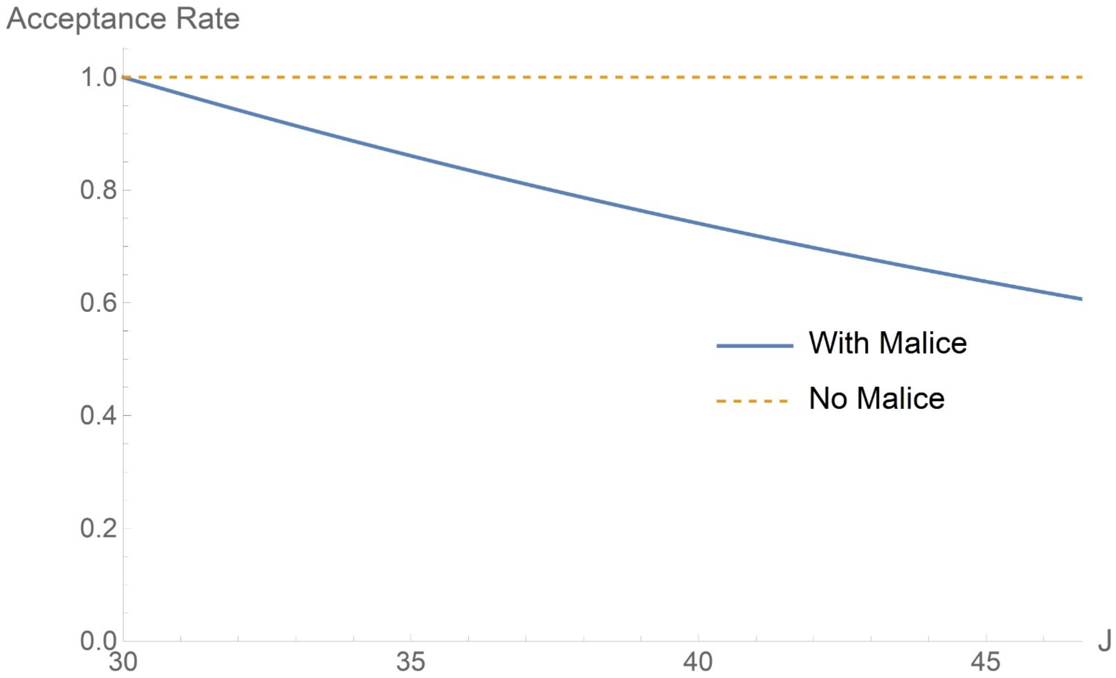


Figure 2: Semi Fully-Revealing Equilibrium

Of course, as the malice factor increases the required cutoff $\hat{J}$ rises, although it must be lower than 80 in this numerical example. Where $\mu = 1$, for instance, the cutoff is $\bar{J} < 50$. We now turn to analyze the case in which $\bar{J} > \hat{J}$

### 5.2. Binary Equilibrium

The foregoing analysis cannot hold where $\bar{J} \geq \hat{J}$. Consider a non-malicious plaintiff $J_i > \hat{J}$. By Lemma 1, this plaintiff is better off reducing her offer rather than proposing a truthful one. By Lemma 2, she prefers to reverse-mimic to a type higher than the lowest one and offer $S' > T$. But in that case, the defendant is better off always accepting that $S'$ – precisely because the non-malicious type attempts to mask as a weaker type. However, if the defendant always accepts demands higher than $T$, types with lower expected damages would mimic upwards.

Given this risk of mimicking, one sustainable strategy for the defendant is to always reject an offer higher than $T$ (and accept with certainty offers $S = T$)[1]. Observe that, given the strategy of the defendant, stronger types would prefer rejection to acceptance at $S = T$. This leads us to the following proposition:

**Proposition 2.** *Suppose $\bar{J} > \hat{J}$. Then, a perfect Bayesian equilibrium exists such that:*

$$(i)\ S^*(J) = \begin{cases} > T & \text{if } M = 0 \text{ and } J > J'_{NM} \text{ or } M = 1 \text{ and } J > J'_M \\ T & \text{elsewhere.} \end{cases}$$

[1]For a somewhat similar logic that generates a 100% rejection (in an NEV signaling model), see Farmer and Pecorino (2007)

*(ii)* $R^*(S) \longrightarrow \{Accept, Reject\} \begin{cases} (0,1) & for\ S > T \\ (1,0) & elsewhere. \end{cases}$

$$(iii)\ \ B^*(S) = \begin{cases} \beta \int_{c_P}^{J'_M} JdF + (1-\beta) \int_{c_P}^{J'_{NM}} JdF & if\ S = T \\ \beta \int_{J'_M}^{\bar{J}} JdF + (1-\beta) \int_{J'_{NM}}^{\bar{J}} JdF & elsewhere. \end{cases}$$

*Where:* $J'_{NM} = c_P + T$; $J'_M = \frac{c_P + T - c_D\mu}{1+\mu}$; $\hat{J} = c_P + T\frac{\mu}{1+\mu}$

*Proof. i)* First, given the defendant's strategy, a plaintiff must either demand (and obtain) $T$ or opt to get rejected (by making a demand higher than $T$) and go to trial. A malicious plaintiff gains from trial $J(1+\mu) + c_D\mu - c_P$, which is higher than $T$ only if $J > J'_M$. Thus, malicious types above $J'_M$ make a demand higher than $T$ and go to trial, while all $J < J'_M$ demand $T$ and settle with certainty. Non malicious plaintiffs' trial payoff is $J - c_P$, which is higher that $T$ only for $J > J'_{NM}$. Thus, NM plaintiffs higher than this threshold go to court by asking a demand higher than $T$, while those below the threshold propose $T$ (which is always accepted). Proposing $S < T$ is inferior to proposing $T$ (which is always accepted).

*ii)* The defendant rejects demands greater than $T$ and accepts otherwise. The defendant has no incentive to reject a demand of $T$ (or lower), as according to her beliefs demands $T$ mostly come from both malicious and non-malicious types whose expected damages is higher than the lowest type, $c_P$ (against which the defendant's liability at trial is just $T$). The defendant also has no incentive to accept a demand higher than $T$, as she cannot improve her payoff and would induce mimicking by doing so.

■

The dynamics that we have described in the previous subsection, then, lead to a qualitatively different result. The desire of non-malicious types to reverse mimic, that is, to mimic downward, persists (Lemma 1). But due to the inability to sustain a revealing equilibrium beyond the lowest settlement demand, $\underline{S} = T$, the defendant rejects all offers above it. As a result, all plaintiffs, except those with the highest expected payoffs, offer $T$ and settle with certainty, whereas the highest types litigate with certainty — hence we term this state binary equilibrium. More precisely, non-malicious plaintiffs above $J'_{NM} = c_P + T$ and malicious plaintiffs above $J'_M = \frac{c_P+T-c_D\mu}{1+\mu}$ should litigate. Observe that $J'_{NM}$ is necessarily bigger than $J'_M$; intuitively, malicious types are less eager to settle. [2] To demonstrate, in our recurrent numerical example ($\underline{J} = 30$, $c_P = 30$, $c_D = 20$, $\mu = 0.5$), $\hat{J} = 46.67$, $J'_{NM} = 80$ and $J'_M = 46.67$, suggesting that litigation is expected only from non-malicious plaintiffs between 80 and 100 and malicious types between 46.67 and 100.

To the extent that $\bar{J} > \hat{J}$ and we are in the binary equilibrium, the rate of settlements could be higher or lower relative to the baseline case. To illustrate, consider

[2] Observe that $J'_{NM}$ is also bigger than $\hat{J}$. Note that $J'_M$ can be higher or lower than $\hat{J}$

again the numerical example in Figure 1, and assume a uniform distribution of plaintiffs types between 30 and 80, and an equal proportion of malicious and non-malicious types. One can verify that in the uni-dimensional case, namely, where spite is common knowledge, the average rate of settlements among non-malicious (malicious) plaintiffs is approximately 63% (52%). Overall, then, the settlement rate is 57.5%. When we move to the bi-dimensional case, the semi fully-revealing is not sustainable, as $\bar{J} = 80 > 46.67 = \hat{J}$. Under the binary equilibrium, all non-malicious and 1/3 of the malicious plaintiffs settle ($J'_{NM} = 80$, $J'_M = 46.67$). Assuming equal proportions, the overall acceptance rate is two thirds, higher than in the benchmark case. Where $\mu$ and $c_D$ are low, $J'_N$ rises towards $J'_{NM}$, and more malicious types settle. Where $c_P$ is high, both cutoffs are higher. Finally, where the distribution is stretched, that is, $\bar{J}$ increases, there is more litigation, as stronger types do not settle. Figure 3 illustrates, using our recurrent numerical example, and assuming $\bar{J} = 120$. One can see that weak plaintiffs always settle and strong plaintiffs always litigate, where the cutoff is lower for malicious (solid blue line) plaintiffs and higher for NM ones (dashed orange line):

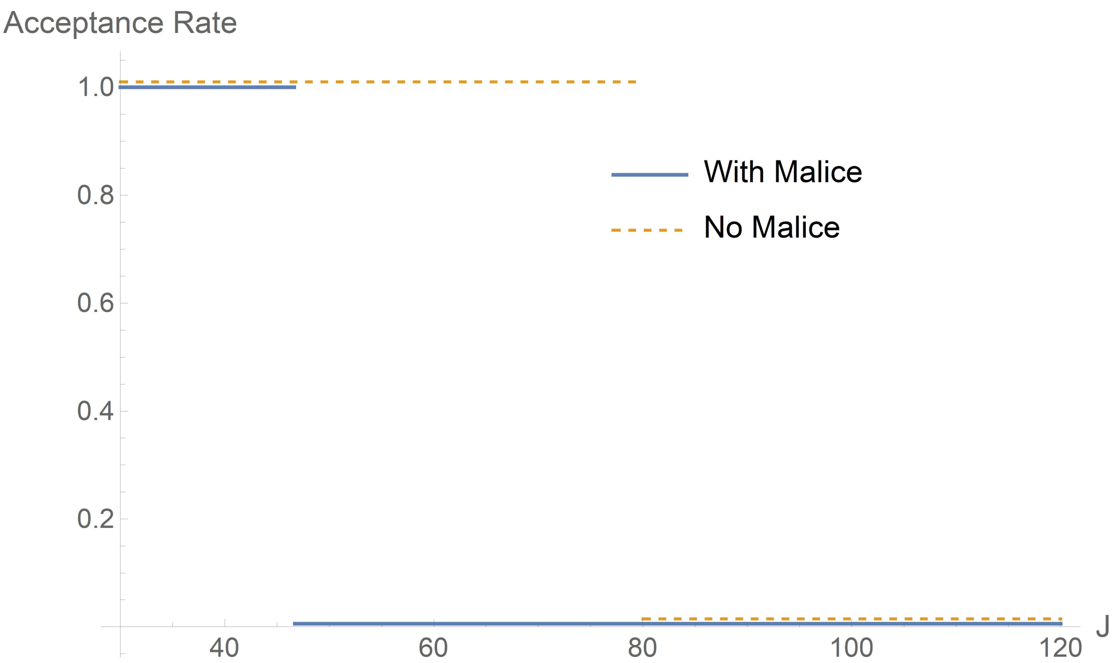


Figure 3: Binary Equilibrium

#### 5.2.1. An Alternative: Semi-Pooling Equilibrium

We have shown that where $\hat{J} < \bar{J}$ a result in which all non-malicious plaintiffs settle and all malicious plaintiffs reveal ("semi fully-revealing") is not sustainable. Rather, the defendant is induced to reject all but the lowest offer — what we termed as "binary equilibrium." Below we briefly show that under certain conditions the dynamics we discuss in the binary equilibrium could give rise to a more complicated equilibrium, in which intermediate plaintiffs with and without malice pool on a certain point between $\underline{J}$ and $\bar{J}$, which we denote $\dot{J}$. As in the binary equilibrium, the defendant again rejects all offers but the lowest one; but here offers that come from revealing $\dot{J}$, $\dot{S} = \dot{J} + c_D$, are accepted by a constant probability $\dot{p}$.

Intuitively, the semi-pooling equilibrium is sustainable where the defendant is just indifferent between accepting and rejecting offers $\dot{S}$. This requires sufficient mimicking from below and reverse-mimicking from above $\dot{J}$, to maintain the defendant's indifference (cf., Lavie and Tabbach (2023), pp. 203-208, who define a similar equilibrium in another context). Accordingly, in the equilibrium we construct, below $\dot{J}$, all non-malicious types settle for $\underline{S}$ and all malicious types mimic and offer $\dot{S}$. Plaintiffs higher than $\dot{J}$ are tempted to reverse mimic to $\dot{J}$ and offer $\dot{S}$. The cutoff $\dot{J}$ is chosen such that the average type who mimics(or reverse mimics) to $\dot{J}$ just equals $\dot{J}$ — hence the defendant is indifferent between accepting and rejecting.

First, consider non-malicious plaintiffs. If a NM plaintiff of type $J_i$ chooses to mimic to $\dot{J}$, her payoff is:

$$\pi_{NM}(J_i, \dot{S}) = \dot{p} * \dot{S} + (1 - \dot{p})(J_i - c_P), \tag{6}$$

which rises with $J$. This suggest that where $\pi_{NM}(\dot{J}, \dot{S}) = \underline{S} = T$, all non-malicious plaintiffs between $\underline{J}$ and $\dot{J}$ prefer to reverse mimic to $\underline{S} = T$ with certain acceptance (and $\dot{J}$ is just indifferent between proposing $T$ and $\dot{S}$). Accordingly, one can verify that the following acceptance rate drives all NM plaintiffs below $\dot{J}$ to propose the lowest offer:

$$\dot{p} = 1 - \frac{\dot{J} - c_P}{T}. \tag{7}$$

Observe that $\dot{p}$ is decreasing with $\dot{J}$. Intuitively, as the settlement $\dot{S}$ increases, one needs a lower acceptance rate $\dot{p}$ to convince low non-malicious types to prefer full acceptant of $T$ over the risky $\dot{S}$.

Recall that in the semi-pooling equilibrium the defendant rejects any offer higher than $\dot{S}$. Equations 6 and 7 thus suggest that non-malicious plaintiffs just higher than $\dot{J}$ will prefer to reverse-mimic and offer $\dot{S}$ (accepted with $\dot{p}$) over proposing $\underline{S} = T$ (accepted with certainty). But as $J$ increases, the option of trial, which yields $J - c_P$ looms more attractive. We will denote $J''_{NM}$ the highest non-malicious $J$ that (weakly) prefers reverse-mimicking to $\dot{S}$ over trial. By equating $\pi_{NM}(J''_{NM}, \dot{S})$ and $J''_{NM} - c_P$, one can verify that $J''_{NM} = \dot{J} + T$.[3] Hence, non-malicious types in the range $[\underline{J}, \dot{J})$ reverse-mimic to $\underline{J}$; non-malicious types in the range $(\dot{J}, min\{J''_{NM}, \bar{J}\}]$ reverse-mimic to $\dot{J}$; and higher types (if they exist) prefer rejection and a certain trial.

We can now turn to analyze the behavior of malicious plaintiffs. Similarly to Equation 6, we can characterize the payoff of a malicious plaintiff of type $J_i$ who chooses to mimic to $\dot{J}$:

$$\pi_M(J_i, \dot{S}) = \dot{p} * \dot{S}(1 + \mu) + (1 - \dot{p})(J_i(1 + \mu) + c_D * \mu - c_P). \tag{8}$$

As before, $\pi_M(J_i, \dot{S})$ increases with $J_i$. This suggests that if the lowest malicious plaintiff, $\underline{J} = c_P$, prefers mimicking to $\dot{J}$ over settling for sure for the lowest offer $T$, then higher malicious plaintiffs will likewise prefer mimicking to $\dot{J}$. One can verify

[3]Observe also that $J''_{NM} = \dot{J} + T > c_P + T > c_p + \frac{\mu T}{1+\mu} = \hat{J}$. And, as we will discuss below, by construction $\hat{J} > \dot{J}$ to enable the semi-pooling equilibrium.

that where $\hat{J} > \dot{J}$, indeed all malicious types prefer mimicking to $\dot{J}$ over proposing $T$ and settling for sure.[4]

However, as before, very strong plaintiffs should prefer rejection (and a certain trial) to mimicking to $\dot{J}$. This happens where $J(1+\mu) - c_P + c_D\mu > \pi_M(J, \dot{S})$. One can verify that the cutoff $J$ that supports this inequality is $J''_M = \dot{J} + \frac{T}{1+\mu}$. Alternatively put, assuming $\hat{J} > \dot{J}$, all malicious plaintiffs from $\underline{J}$ to $min\{J''_M, \bar{J}\}$ pool on $\dot{J}$, whereas stronger plaintiffs (if they exist) prefer certain rejection.

Finally, observe that the defendant is just indifferent between accepting and rejecting $\dot{S} = \dot{J} + c_D$ where the expected value of plaintiffs who offer $\dot{J}$ (weighted by the proportions of malicious and non-malicious) equals $\dot{J}$. We accordingly define $\dot{J}$:

$$\dot{J} = \frac{\beta \int_{c_P}^{min\{J''_M, \bar{J}\}} JdF + (1-\beta) \int_{\dot{J}}^{min\{J''_{NM}, \bar{J}\}} JdF}{\beta F(min\{J''_M, \bar{J}\}) + (1-\beta)(F(min\{J''_{NM}, \bar{J}\}) - F(\dot{J}))} \tag{9}$$

We briefly discuss and demonstrate the semi-pooling equilibrium below (Subsection 8.2 in the Appendix provides full characterization and proof). The semi-pooling equilibrium is somewhat similar in spirit to the binary equilibrium. As in the binary equilibrium, settlement proposals are either accepted or rejected. In addition to these two options, the semi-pooling equilibrium also features a third option, accepting at a constant rate, $\dot{p}$. These options essentially create two "focal points" of settlements. One (as in the previous equilibria) at $\underline{S} = T$, and the other at $\dot{S} = \dot{J} + c_D$. All weak non-malicious plaintiffs (up to $\dot{J}$), as in the two previous equilibria, reverse mimic to the lowest type and offer $\underline{S}$. All intermediate non-malicious (above $\dot{J}$ and up to $min\{J''_{NM}, \bar{J}\}$) reverse mimic to $\dot{J}$); they pool at $\dot{J}$ with weak and intermediate malicious types (from $\underline{J}$ to $min\{J''_M, \bar{J}\}$). Stronger types, if they exist, prefer to go to trial.

The following figure, Figure 4, utilizes our recurrent numerical to show the semi-pooling equilibrium, where $\bar{J} = 58$; we also assume that types are distributed uniformly, and that $\beta = 0.5$, that is, equal proportion of malicious and non-malicious. As in previous figures, the acceptance rates of malicious (blue) and NM (dashed orange) are represented by the y-axis, where the x-axis is the type:

[4] Observe that the lowest type prefers mimicking to $\dot{J}$ where $\pi_M(c_P, \dot{S}) = \dot{p}\dot{S}(1+\mu) + \mu T(1-\dot{p}) > T(1+\mu)$. Or, by dividing in $1+\mu$, this condition becomes $\dot{p}\dot{S} + \frac{\mu T(1-\dot{p})}{(1+\mu)} > T$. But we know that $T = \pi_{NM}(\dot{J}, \dot{S}) = \dot{p}\dot{S} + (1-\dot{p})(\dot{J} - c_P)$ (we derived Equation 7 by this equation). Hence, the condition is $\dot{p}\dot{S} + \frac{\mu T(1-\dot{p})}{(1+\mu)} > T = \dot{p}\dot{S} + (1-\dot{p})(\dot{J} - c_P)$, which yields $\dot{J} < c_P + \frac{\mu T}{1+\mu} = \hat{J}$.

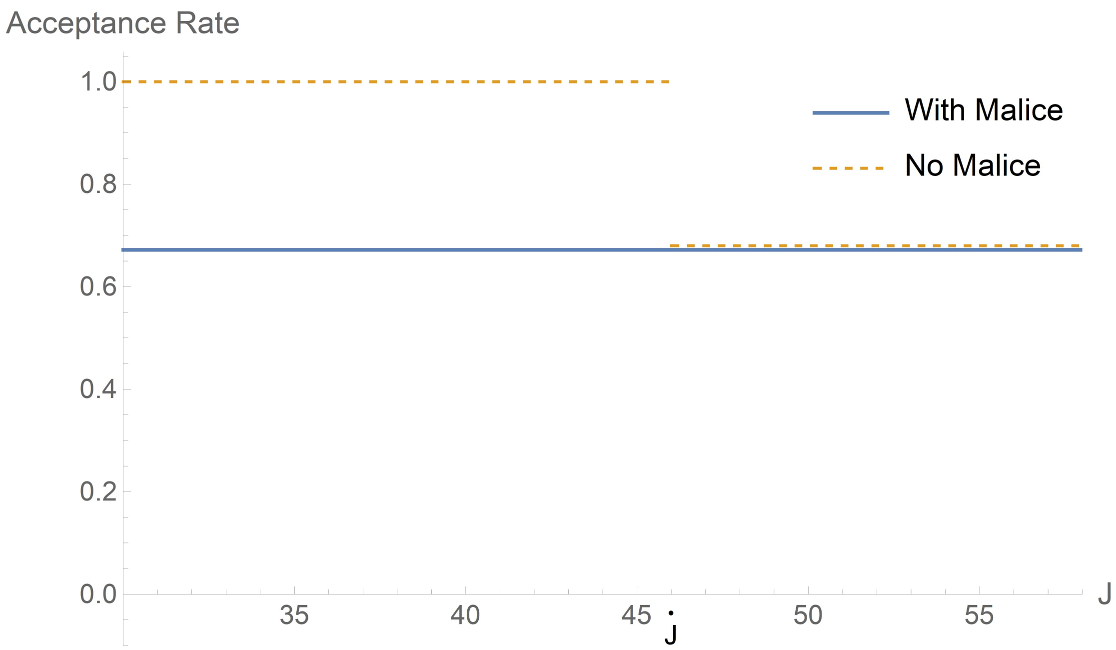


Figure 4: Semi Pooling Equilibrium

Under this example, we can calculate $\dot{J}$, which approximately equals 46.4 (less than $\hat{J} = 46.67$).[5] One can see the two “focal” settlements in this case. The first, $\underset{\sim}{S} = 50$, attracts all non-malicious plaintiffs below $\dot{J} = 46.4$. The defendant always accepts this offer. The second, $\dot{S} = \dot{J} + c_D = 66.4$, attracts all malicious plaintiffs as well as strong non-malicious plaintiffs, above $\dot{J} = 46.4$. The defendant accepts this offer with $\dot{p} \approx 0.67$. In this example there are no strong plaintiffs who opt for trial.

A critical enabling assumption for the semi-pooling equilibrium is the assumption that $\hat{J} > \dot{J}$, which ensures that all malicious plaintiffs prefer to propose $\dot{S}$ over $\underset{\sim}{S} = T$ (footnote 3 and accompanying text). This assumption, then, ensures that the semi-pooling equilibrium exists. In the numerical example that underlies Figure 4, for instance, stretching $\bar{J}$ beyond $\bar{J} = 58$ undermines this conditions. A higher malice factor increases $\hat{J}$ and decreases $J''_M$ and thus facilitates the semi-pooling equilibrium.[6] $\dot{J}$, of course, depends on the distribution. Left-skewed distributions, for example, allow for larger $\dot{J}$.

We stress that we do not claim uniqueness here. There might be other possible equilibria. For instance, one can further extend the logic to create additional “focal points” of settlements where $\bar{J}$ increases. Likewise, even where $\hat{J} < \dot{J}$ one can possibly find a settlement that leaves the defendant just indifferent (and again, that point is distribution dependent).

[5]The cutoffs $J''M$ and $J''NM$ are approximately 80 and 95 — well above $\bar{J}$

[6]To demonstrate, where $\mu$ increases to 1, $\hat{J} = 55$, and $\bar{J}$ could increase to 72 and still leave $\dot{J}$ below $\bar{J}$. In that case, as in the example in Figure 4, there are no stronger plaintiffs who prefer to litigate

## 6. Discussion

The following summarizes, compares, and discusses our results. We have seen that the introduction of additional informational gaps, through another dimension of private information, leads to interesting and sometimes unexpected results.

*Settlement Rate.* The first parameter of interest is the rate of settlements. By and large, we have found that the introduction of an additional dimension of private information encourages (perhaps counterintuitively) settlements, provided that the distribution of plaintiff types is not too wide. If the malice factor and trial expenses are sufficiently large with respect to $\bar{J}$, the strongest plaintiffs, the parties reach the semi fully-revealing equilibrium. In this equilibrium all non-malicious plaintiffs settle, and the malicious ones settle according to $pM(S)$, the benchmark acceptance rate. Where $\bar{J}$ increases, we shift to the binary equilibrium, in which weaker plaintiffs (with or without spite) always settle and strong plaintiffs always litigate. Under certain conditions (and sufficient malice relative to other parameters), we found that there exists a semi-pooling equilibrium, in which the weakest NM plaintiffs always settle and strongest plaintiffs, both malicious and non-malicious, always litigate (as before); but, intermediate plaintiffs pool on intermediate offer (with intermediate acceptance rate).

Although these differences make comparisons difficult, we can stress the following themes. (i). Unambiguously, the weakest non-malicious plaintiffs reverse-mimic to the lowest offer and settle with certainty. (ii), malice, perhaps counterintuitively, in principle encourages settlement, as it aggravates the condition of non-malicious plaintiffs and facilitates their tendency to reverse-mimic. However, (iii) above a certain cutoff, the strongest plaintiffs (if they exist) always litigate, where malicious plaintiffs are more eager to do so. Thus, when we stretch the distribution (higher $\bar{J}$), we should expect more litigation relative to the benchmark case. Therefore, malice is not necessarily correlated with more/fewer trials.[7]

*Settlement and Merits.* A more definite prediction concerns the merits of settlements. As we have shown, we have proved in all models that at least some weak, non-malicious plaintiffs desire to "reverse-mimic" to lower types due to the inability of the defendant to decipher the second dimension.[8] Under certain conditions, all non-malicious plaintiffs reverse-mimic to the lowest type (the semi fully-revealing case). Under other conditions, intermediate non-malicious plaintiffs reverse-mimic to a certain, intermediate point (semi-pooling). The behavior of malicious types is more nuanced. In the semi-fully-revealing case they always reveal. But in the other cases, weak malicious plaintiffs pull back to the lowest type (binary case); or, intermediate malicious plaintiff pool together on an intermediate point, through mimicking by weaker types and reverse-mimicking by intermediate ones (semi pooling case).[9]

More generally, our results suggest that the introduction of a second dimension of

[7]The picture is quite similar with respect to the two-type model, where the introduction of an additional dimension of private information increases the rate of trials initially, but as malice becomes larger (again, relative to the strongest plaintiffs), more settlements are expected.

[8]This is true in the continuous case. In the two type model, non-malicious plaintiffs pull back where the malice factor is sufficiently high

[9]In the two-type model we observe a similar phenomenon where weak malicious types may mimic upwards (and pool with the strong, non-malicious types who reverse mimic).

private information breaks the familiar fully-revealing equilibrium (Reinganum and Wilde (1986)). Instead, settlements are discontinuous. One should expect "focal points" of settlements. One, which exists in all models, occurs at the beginning of the distribution, $\underline{S}$. In the semi-pooling case, there is an additional "focal" settlement for intermediate types. These points could influence, then, the interpretation of empirical results — indeed, some experiments find that participants who play a pre-trial bargaining game tend to converge on mid-points (cf., Guerra et al. (2025)).

Relatedly, our results may undermine the monotonicity assumption. It is almost axiomatic that "when the plaintiff has superior information, it is more likely to litigate stronger cases" (Klerman and Lee (2014), p. 212). Accordingly, one can infer from litigated cases, selection effects notwithstanding (id.). The introduction of an additional dimension of private information could flip this prediction. Indeed, where the distribution of malicious and non-malicious plaintiffs is identical, strong plaintiffs are still more likely to litigate. But we can easily construct a contrary example, with different distributions. Consider, for instance, the dynamics that Figure 2 describes. All non-malicious plaintiffs settle, and malicious plaintiffs tend to litigate more where their cases are stronger. If non-malicious types have better cases, overall we may see less litigation for stronger types.

*Over-Rejection and Overly Generous Offers.* A recurrent point that emerges is the over-rejection of "revealing" offers. This point underlies our discussion. As the uninformed party cannot decipher malice, she will tend to reject offers (compared to the benchmark case in which there are only non-malicious types). Expecting this behavior, the non-malicious types will tender a more generous offer than they "should." Alternatively put, our results suggest that in many cases settlements are lower than the corresponding merits of the proposer — even where the proposer has all the bargaining power (and the receiver behaves rationally).

Again, this point bears empirical predictions. Experimental literature on pre-trial bargaining has shown that participants tend to over-reject (where they play the receiver) on the one hand, and leave money on the proverbial table by proposing a generous offer, on the other hand (for an example see Pecorino and Van Boening (2018), who assess the proportion of money left as $\frac{1}{6}$ of the surplus). Of course, these phenomena are often attributed to behavioral motivations. However, our analysis could rationalize these behaviors through the additional dimension of private information.[10]

*Payoffs.* Another recurrent finding concerns payoffs. Although there is another dimension of private information, the uninformed party in our setup, the defendant, is generally better off.[11] The intuition is as follows. The existence of (unobservable) malicious types allows the defendant to commit to be more aggressive towards non-malicious types in rejecting settlement proposals. The defendant's behavior pushes the non-malicious types to decrease their offer — a clear gain for the defendant.

[10]Interestingly, Guerra et al. (2025) find that, when played against a robot, participants playing receivers virtually always accepted zero-surplus offers from weakest types, contrary to behavioral predictions; but nonetheless still over-rejected zero-surplus offers from strong types. These behaviors fit our predictions

[11]The defendant is always better off under the continuous model, but this is not necessarily true in the two-type model. Particularly, where weak types are over-represented in that model, the defendant may be worse off.

The non-malicious plaintiffs present the mirror-image phenomenon — in principle, they are always worse off. They are, by and large, worse off, as they have to deviate from the standard equilibrium and decrease their offer. In general, then, we can say that non-malicious plaintiffs subsidize the uninformed defendant, as the latter cannot tell whether they are motivated by spite and more eager to litigate. We note that a similar dynamics could occur with respect to malicious plaintiffs. Due to the co-existence of spiteful and non-malicious plaintiffs and the risk of mimicking, the defendant could be committed to always reject certain offers. This reduces the payoff of at least some of the malicious plaintiffs, for instance, the strongest types in the binary and semi-pooling states, who are now bound to always go to trial without extracting any settlement surplus.[12]

*Conveying Private Information.* As non-malicious types are worse off, one wonders whether they could somehow convey information regarding their "benevolent" motivations, other than reducing their offer (compare Lavie and Tabbach (2020)). We suspect that non-malicious types do not have a credible way to signal their innocent motivations (where, on the other hand, spiteful plaintiffs can mask their true motive).

*TIOLI Assumption.* We have assumed throughout that the informed type proposes a take-it-or-leave-it (TIOLI) settlement offer, which allows that type to extract the entire settlement surplus. One wonders whether a more equal allocation of the surplus eliminates our results. In that case, the second dimension should be reflected — at least partially — in the settlement offer. We believe that the general dynamics that our results depict persists even where the settlement reflects part of the gain to the informed type from the second dimension. Such a setup would still result in informed types that differ with respect to the gain they have from settlements, and thus, as before, informed types who propose a similar settlement but require a different acceptance function to maintain equilibrium.

We can illustrate our point by the following example. Suppose our recurrent numerical example, $c_P = 30$, $c_D = 20$, and consider a non-malicious plaintiff of type $J_{NM} = 100$. In case of trial, the defendant is bound to pay 120, and the plaintiff net gain is 70 (indeed, the joint trial costs $T$ are 50). Assuming that the parties always settle at the middle, the relevant, revealing settlement is $S_{NM} = \frac{120+70}{2} = 95$.

Now assume a malice factor $\mu = \frac{1}{2}$ and consider a malicious plaintiff of type $J_M = 76$. In case of trial against that plaintiff, the defendant pays 96, inclusive the defendant's trial expenses. Given the 0.5 malice factor, the gain of $J_M = 76$ at trial is $76*1.5+20*0.5-30 = 94$. By our assumption that the parties settle halfway, a revealing settlement from this type equals $S_M = \frac{95+94}{2} = 95$. We have shown, then, that two different plaintiffs — malicious plaintiff with expected damages of 76 and non-malicious plaintiff with expected damages of 100 — tender the same revealing settlement offer. We can of course generalize this example. By the logic of this paper, the defendant will take the offer to trial as if it comes from malicious types, to prevent mimicking ($p_M(S)$ rather than $p(S)$. The more parsimonious rate of acceptance should again push the non-malicious type to mimic as a lower type (and enjoy the higher acceptance rate).

By contrast, this dynamics does not exist where all the gain of the informed type

[12] In the semi-pooling state, this dynamics also lead the intermediate malicious types to pool on an intermediate settlement, suggesting that some of them are better and others are worse off.

from the second dimension is reflected in the settlement offer. The screening model, where the uninformed has the TIOLI power, illustrates. In that model, the offer by the uninformed attempts to attract the entire settlement surplus, including all the gains from trial that the informed plaintiff expects. Alternatively put, the two dimensions collapse into a single one.

*Alternative Examples of Private Information.* We have used malice as a primary example that illustrates the second dimension, affecting the informed's gain without affecting the uninformed's loss at trial. This is, of course, just an example. We can think of various other examples of information that could be private information for one of the parties, but have no effect on the expected judgment at trial. A straightforward example is plaintiffs that differ on (dimension 1): expected damages; and (dimension 2): their trial expenses, that is, for some plaintiffs trial is cheaper, such that these plaintiffs have a greater "appetite" for trial, although their expected judgment at trial is independent of their trial costs.

Indeed, we can easily construct our two type example along these lines. Recall that in our expository, two-type numerical example (Section 2) the expected damages of strong and weak plaintiffs are respectively 55 and 30; and the costs of trial are 30 and 20 (for plaintiffs and defendants, respectively). we have shown in Section 2 above that $p = \frac{2}{3}$ maintains a revealing equilibrium.

Now consider a plaintiff whose costs of trial are lower by a factor $\mu$. We may think of plaintiffs who have an easy access to lawyers as a straightforward explanation for this variety in the plaintiffs' population.[13] Thus, if this so-called "privileged" plaintiff goes to trial, her trial expenses are just $30\mu$. With only "privileged" plaintiffs, and assuming $\mu = 0.5$, one can verify that the fully-revealing rate of acceptance is $0.58\frac{1}{3}$, lower than the base acceptance rate, $\frac{2}{3}$.[14] Of course, the acceptance rate decreases with $\mu$; for instance, if $\mu = \frac{1}{6}$, the acceptance rate of privileged, strong plaintiffs in equilibrium is just $\frac{1}{2}$. Intuitively, as in the malice examples before, privileged plaintiffs are more eager to go to trial, hence the rejection rate should be higher in equilibrium.

We can now evaluate when and whether reverse mimicking exists in this example. Intuitively, the lower is $\mu$, privileged plaintiffs have a greater appetite for trial, reducing the acceptance rate of offers of 75. This, in turn, aggravates the plight of the non-privileged, who has to suffer a lower acceptance rate by virtue of the defendant's inability to observe the second dimension, costs of trial. A non-privileged plaintiff's gain from offering 75 is $p * 75 + (1 - p)(55 - 30)$. Where $\mu = 0.5$, we have shown that $p = 0.58\frac{1}{3}$, hence the value of offering 75 is $\approx 54.16$, which is still higher than the alternative of offering 50 and settling for sure. But one can verify that where $\mu = \frac{1}{6}$, the strong non-privileged type (weakly) prefers to reverse-mimic and proposes a low offer of 50.

We can, of course, further generalize this example. We conclude that the dynamics that we highlight in this paper extends to other contexts in which there are two di-

[13]We can think of plenty other real-world reasons for this difference between plaintiffs, which is their private information. Plaintiffs may have legal knowledge and/or experience, enabling them to better monitor their lawyer (and cut legal expenses). Some plaintiffs may simply be more litigious, with lower non-monetary costs of trial.

[14]This rate is derived by equating 50, the value of revealing as a weak type and settling with certainty, to $p * 75 + (1 - p)(30 - 30 * \mu)$, that is, the value of going to trial for privileged plaintiffs.

mensions of private information — one that pertains to the expected judgment, and the other that affects the payoff of the informed party from trial (without directly affecting the outcome at trial).

## 7. Conclusion

While traditional models of litigation assume private information in a single dimension, we explore the effects of adding a second dimension of private information for the same party, such that this dimension affects the informed type's appetite for trial, without affecting the expected judgment at trial. We focus on the specific example of malicious motivations; plaintiffs differ not only in their expected judgments, but also in the presence or absence of malice, and both dimensions are unknown to the defendant.

Introducing this second dimension leads to surprising effects. First, the familiar fully-revealing result is not sustainable with bi-dimensional private information. Those who have weaker "appetite" for trial tend to "reverse mimic" to the weakest type (and settle with certainty). This effect may increase the rate of settlements. Second, the uninformed party is counterintuitively better off — intuitively, the addition of private information in the second dimension allows her to be more aggressive, which facilitates more generous offers. Third, bi-dimensional private information gives rise to various equilibria, semi-fully-revealing, binary (accept/reject), and semi-pooling ones.

Our results bear empirical predictions. We predict "focal" settlement points, especially at the low-end of the distribution of types; challenge the assumption regarding monotonicity in settlement rates; and offer a rational explanation for over-rejection and overly generous offers observed in experiments.

# 8. Appendix

## 8.1. The Two-Type Case

This Section discusses more formally the two type case that we presented numerically (Section 2).

*Setup.* We assume that the plaintiff is privately informed concerning two dimensions: her expected damages $J \in \{L, H\}$ and her malicious motivations $M \in \{0, 1\}$. As to the first dimension, there are two types of plaintiffs, with lower and higher expected damages, $L$ and $H$, respectively. The defendant only knows the share of each group, $\alpha$ and $1 - \alpha$, respectively. The second dimension relates to the inner motivations of the plaintiffs. We assume that a share $\beta$ of the plaintiffs are motivated by malicious motivations, whereas the remaining plaintiffs are non-malicious (NM). In that case, in addition to their monetary payoff, these malicious types enjoy a fraction $\mu$ of the rival's loss (from trial or settlement). A higher $\mu$ reflects a more significant spite from the plaintiff's side, and a larger gain from the rival's loss. Thus, if the defendant pays at trial $H$, the plaintiff profits $H(1 + \mu)$. Again, the inner motivation of the plaintiff is her private information. The defendant only knows that a share $\beta$ of the plaintiffs have malicious motivations. We also assume that these two dimensions — damages and malice — are independent of each other.

The sequence of the game is as follows. At the first stage the plaintiff proposes a single take-it-or-leave-it offer $S$. The defendant responds by either accepting or rejecting ($R : (S) \to \{Accept, Reject\}$). If the offer is accepted, the payoffs are determined by the settlement offer: the defendant loses $S$, non-malicious (NM) plaintiffs gain $S$, and malicious plaintiffs gain $S(1+\mu)$. In case the offer $S$ is rejected and the case goes to trial, the trial expenses of each party are $c_D$ for the defendant and $c_P$ for each plaintiff's type. We denote the sum of trial costs as $T = c_P + c_D$. If trial occurs, the payments are in accordance with the first dimension, that is, $H$ or $L$. The malicious plaintiff enjoys an additional profit of $\mu$ times the judgment, reflecting her malice towards the defendant. In addition to the augmented benefit from trial, the malicious type also gains from the trial costs that her rival expended, by virtue of her spite towards her rival. This benefit equals $\mu C_D$.

*Benchmark Case: Uni-Dimensional Case.* Suppose that there are no plaintiffs with malicious motivations. We could thus expect a fully-revealing equilibrium. In this familiar result all plaintiffs reveal their type. They do so by offering a settlement offer that reflects the defendant's expected costs from going to trial, namely, $S_L = L + c_D$ for low-damages plaintiffs and $S_H = H + c_D$ for high-damages plaintiffs. Low offers $S_L$ are accepted by the defendant with certainty in equilibrium. To prevent mimicking by $L$ types, offers $S_H$ are accepted by the defendant in equilibrium such that $L$ types are just indifferent between revealing (and earning $S_L$ with certainty) and attempting at mimicking (and earning $S_H$) but risking a trial and paying $S_L - c_P$ if their offer is rejected. One can verify that the acceptance rate that supports equilibrium, denoted $p_0$, is:

$$p = \frac{T}{T + H - L}. \tag{10}$$

Observe that in this fully-revealing equilibrium the defendant is indifferent between going to trial and settling (the plaintiff extracts all surplus).

We can now sketch the parallel setup, where all plaintiffs have malicious motivations. We again face a fully-revealing equilibrium, with the same offers $S_L$, $S_H$ — observe that the malice factor does not change the expected liability of the defendant from trial. However, malice adds to the appeal of trial, and, by extension, to the desire of $L$ types to mimic. A malicious plaintiff of type $L$ expects $S * (1 + \mu)$ if her offer is accepted, and $p_M S_H(1+\mu)+(1-p_M)((1+\mu)L-c_P+\mu c_D)$ in case of trial. One can verify that the rate at which the defendant has to take $S_H$ offers to trial where all plaintiffs are motivated by spite, denoted $p_M$ is:

$$p_M = \frac{T}{T + (1 + \mu)(H - L)}. \tag{11}$$

Observe that $p_M$ decreases as the malice factor $\mu$ increases, and is necessarily lower than the acceptance rate without malice, $p_M < p_0$. Intuitively, trial looms more attractive as spite increases, which requires a higher rate of rejection to prevent mimicking. As before, in this fully-revealing equilibrium the defendant is indifferent between going to trial and settling, regardless of the rate at which it rejects high offers.

*Two-Dimensional Case.* We now consider private information in two dimensions: damages and malice. Given the binary nature of each dimension, the result is a four-type setting. We will denote non-malicious types as $H_0$ and $L_0$, and malicious ones as $H_M$ and $L_M$.

First, observe that a revealing offer from both $L$ types equals $S_L$, regardless of their malice (as malice has no effect on the defendant's loss at trial). Hence, both $L$ types reveal and offer $S_L$, which is accepted with certainty.

However, the differences regarding spite do affect the $H$ types. With full information on spite, in equilibrium the defendant would have accepted offers $S_H$ from malicious types with $p_M$ and from non-malicious types with $p_0$, where $p_0 > p_M$. Lacking this information, the defendant is worse off responding to $S_H$ with the more generous rate, $p_0$, as it induces mimicking by malicious $L$ plaintiffs. Hence, the defendant is better off responding to $S_H$ with $p_M$. The payoff of non-malicious H types given $p_M$ is thus:

$$\pi_H(p_M) = p_M S_H + (1 - p_M)(H - c_P). \tag{12}$$

The non-malicious $H$ types, then, lose as a result of the other, malicious $H$ types, and the inability of the defendant to decipher malice in the bi-dimensional case. As a result, the non-malicious $H$ may be tempted to propose a lower offer, or, "reverse-mimic." Suppose, for instance, that the non-malicious $H$ pulls back to the lowest offer, $S_L$, which is accepted with certainty. One can verify that such reverse-mimicking — or, $S_L > \pi_H(p_M)$ — occurs where the malice factor is sufficiently high:

$$\mu \geq \frac{H - L}{T - (H - L)} \equiv \tilde{\mu}. \tag{13}$$

With such a high malice factor, $p_M$ is low and all non-malicious $H$ types are better off proposing the low $S_L$ and pool with $L$ plaintiffs, malicious and non-malicious alike.

But where the non-malicious $H$ are expected to pool with the $L$ types, the latter can now raise their offer beyond $S_L$, and the defendant should still accept it with certainty. Recall that a share $\alpha$ of the population has low expected liability, $L$, and a share $\beta$ has malicious motivations. Hence, a share $1-\beta(1-\alpha)$ of the population pools: $H_0$, $L_0$, and $L_M$ (respectively, $(1-\alpha)(1-\beta)$, $\alpha\beta$, and $\alpha(1-\beta)$). Therefore, the defendant would be willing to accept a higher settlement, where the highest settlement equals his alternative of going to trial against these three types:

$$\tilde{S} = \frac{(1-\alpha)(1-\beta)(H+c_D)+\alpha(L+c_D)}{1-\beta(1-\alpha)}. \tag{14}$$

Indeed, $S_L < \tilde{S} < S_H$. With a higher settlement that is accepted with certainty, the defendant can increase the probability with which she accepts higher offers $S_H$ (such that $L_M$ types do not find it worthwhile to mimic). The new acceptance rate is derived by $L_M$ new indifference function:

$$S(1+\mu) = p(S)S_H(1+\mu) + (1-p(S))((1+\mu)L - c_P + \mu c_D). \tag{15}$$

This yields the following acceptance rate of $S_H$ offers, given that the three types $H_0, L_0, L_M$ pool to $S_L < S < \tilde{S}$:

$$\tilde{p}(S) = \frac{c_P - c_D\mu + (S-L)(1+\mu)}{T+(H-L)(1+\mu)}. \tag{16}$$

This acceptance rate ensures that $L_M$ types will not be tempted to propose a higher offer $S_H$. As our proof shows, it also guarantees that $H_0$ types still prefer to "reverse-mimic," given a sufficiently high malice factor. Observe that $\tilde{p}(S)$ converges to $p_M$ where $\tilde{S} = S_L$, and that $\tilde{p}(S) > p_M$. However, although the lower types (together with $H_0$) propose an offer higher than $S_L$, $\tilde{p}(S)$ may or may not be greater than $p_0$ due to the spite factor $\mu$. This leads us to the following proposition:

**Proposition.** *The following is a Perfect Bayesian Nash Equilibrium.*

*(i)* $S^* = \begin{cases} S_H & \text{for } (J=H \text{ and } M=1) \text{ or } (M=0 \text{ and } \mu \le \tilde{\mu}) \\ S_L & \text{for } (J=L \text{ and } \mu < \tilde{\mu}) \\ \tilde{S} & \text{elsewhere.} \end{cases}$

*(ii)* $R^*(S) \longrightarrow \{\text{Accept, Reject}\} \begin{cases} (p_M, 1-p_M) & \text{for } S = S_H \text{ and } \mu \le \tilde{\mu} \\ (1,0) & \text{for } S \le S_L \text{ and } \mu \le \tilde{\mu} \\ (\tilde{p}(\tilde{S}), 1-\tilde{p}(\tilde{S})) & \text{for } S = S_H \text{ and } \mu > \tilde{\mu} \\ (1,0) & \text{for } S \le \tilde{S} \text{ and } \mu > \tilde{\mu} \\ (0,1) & \text{elsewhere.} \end{cases}$

*(iii)* $B^*(S) \longrightarrow \{J_H, J_L\} \begin{cases} (1,0) & \text{for } S \ge S_H \\ (\frac{(1-\beta)(1-\alpha)}{\alpha+(1-\beta)(1-\alpha)}, \frac{\alpha}{\alpha+(1-\beta)(1-\alpha)}) & \text{for } S < S_H \text{ and } \mu > \tilde{\mu} \\ (0,1) & \text{elsewhere.} \end{cases}$

Where: $S_H = H + c_D$; $S_L = L + c_D$; $p_M = \frac{T}{T+(1+\mu)(H-L)}$; $\tilde{\mu} = \frac{H-L}{T-(H-L)}$;

$$\tilde{S} = \frac{(1-\alpha)(1-\beta)(H+c_D)+\alpha(L+c_D)}{1-\beta(1-\alpha)};\ \tilde{p}(S) = \frac{c_P-c_D\mu+(S-L)(1+\mu)}{T+(H-L)(1+\mu)}.$$ [15]

*Proof.* **Part I: $\mu \leq \tilde{\mu}$:**

*a)* $L_M$: this type weakly prefers certain $S_L$ settlement to acceptance of $S_H$ with probability $p_M$ (Equation 15). Likewise, it is strictly worse off where the defendant always rejects (offers $S_L < S < S_H$, $S_H < S$). Where the defendant always accepts $S \leq S_L$, the malicious $L$ maximize by offering $S_L$. Hence, offering $S_L$ is a (not unique) best-response.

*b)* $L_0$: For similar reasons, this type prefers certain $S_L$ settlement to acceptance of $S_H$ with probability $p_M$ (this follows from Equation 15 and the fact that $p_M < p_0$). Likewise, where the defendant always accepts ($S \leq S_L$), this type maximize by offering $S_L$.

*c)* $H_M$: This type prefers acceptance of offers $S_H$ with $p_M$ to certain acceptance of $S \leq S_L$ (this follows from the indifference function of the malicious $L$, Equation 15). Other offers, and in particular offers higher than $S_H$, are always rejected. Hence, offering $S_H$ is a best-response.

*d)* $H_0$: By Equation 13, this type is at least weakly better off by offering $S_H$ which will be accepted by $p_M$, than certain acceptance of $S \leq S_L$. Then, proposing less than $S_L$ cannot be a best-response. Other offers, in particular offers higher than $S_H$, are always rejected. Hence, offering $S_H$ is a (perhaps not unique) best-response.

*e)* Defendant: According to her beliefs, offers $S_H$ are attributed to $H$ types. As her alternative of going to trial yields an amount $H + c_D = S_H$, accepting with $p_M$ is a best-response (albeit not uniquely so). Likewise, offers $S_L$ are attributed to $L$ types. As the settlement offer equals the alternative of going to trial, accepting with certainty is a best-response (albeit not uniquely so).

**Part II: $\mu > \tilde{\mu}$:**

*a)* $L_M$: The defendant accepts with certainty offers $S \leq \tilde{S}$ and rejects all other offers except $S_H$ (which is accepted with $\tilde{p}(\tilde{S})$). Hence, $L_M$ is surely worse off proposing any other offer except $S = S_H$ or $S = \tilde{S}$. It follows from Equations 15 and 16 that $L_M$ is indifferent between these two options, and thus she is weakly better off proposing $S = \tilde{S}$ (which is accepted with $\tilde{p}(\tilde{S})$).

*b)* $L_0$: Again, the defendant accepts with certainty offers $S \leq \tilde{S}$, and rejects all other offers except $S_H$, which is accepted with $\tilde{p}(\tilde{S})$. Hence, we can rule out any offer except $S = S_H$ or $S = \tilde{S}$. To see that $L_0$ is strictly better off when proposing $\tilde{S}$, consider the payoff of $L_0$ when she proposes $S_H$:

$$\pi_{L_0}(S_H, \tilde{p}(S)) = \tilde{p}(S)(H + c_D) + (1 - \tilde{p}(S))(L - c_P). \tag{17}$$

One can verify that, given the definition of $\tilde{p}(S)$, this payoff decreases with $\mu$.[16]

[15] We note that the defendant's beliefs concerning the plaintiff's malice are irrelevant to her response, and thus we have not discussed these beliefs.

[16] The derivative $\frac{\partial \pi_{L_0}(S_H, \tilde{p}(S))}{\partial \mu} = -\frac{T(T+H-L)(H+c_D-S)}{(T+(H-L)(1+\mu))^2}$, which is necessarily negative given $H + c_D > S$.

And where $\mu = \tilde{\mu}$, this expressions equals:

$$S + \frac{(H-L)\,(S-(H+c_D))}{T},$$

which is necessarily lower than $S$ for any $S < H + c_D$, including $\tilde{S}$. Hence, $L_0$ is strictly better off proposing $\tilde{S}$.

*c*) $H_M$: For similar reasons, it suffices to show that this type prefers acceptance of offers $S_H$ with $\tilde{p}(\tilde{S})$ to certain acceptance of $S \leq S_L$. This preference follows from the last paragraph and the indifference function of the malicious $L$, Equation 15). Hence, offering $S_H$ is a best-response.

*d*) $H_0$: Similarly to the case of $L_0$ above, it suffices to show that this type (at least weakly) prefers to "reverse-mimic," propose $\tilde{S}$, and settle with certainty; to revealing and offer $S_H$. Similarly to Equation 17, consider the payoff of $S_0$ when she proposes $S_H$ and risk a trial:

$$\pi_{S_0}(S_H, \tilde{p}(S)) = \tilde{p}(S)(H + c_D) + (1 - \tilde{p}(S))(H - c_P). \tag{18}$$

Again, one can verify that this payoff decreases with $\mu$. And where $\mu = \tilde{\mu}$, this simply reduces to $S$. Hence, $H_0$ is weakly better off proposing $\tilde{S}$ where $\mu = \tilde{\mu}$, and is strictly better off proposing $\tilde{S}$ where $\mu$ is greater than the cutoff $\tilde{\mu}$.

*e*) Defendant: Per her beliefs, offers $S \geq S_H$ are attributed to $H$ types. As her alternative of going to trial yields an amount $H + c_D = S_H$, accepting with $\tilde{p}(\tilde{S})$ offers $S_H$ is a best-response (albeit not uniquely so). With these beliefs, rejecting with certainty offers higher than $S_H$ is likewise a best-response. Settlement offers below $S_H$ are attributed to the pool of the three types, $L_M, L_0, H_0$. Equation 14 implies that the defendant is strictly better off accepting offers below $\tilde{S}$; and rejecting (and going to trial) offers $\tilde{S} < S < S_H$. Likewise, the defendant is indifferent between trial and $\tilde{S}$. Hence, accepting (rejecting) for $S \leq \tilde{S}$ ($\tilde{S} < S < S_H$) is a best-response (though not always uniquely so).

■

The foregoing equilibrium has essentially two versions: where the malice factor is above or beyond the cutoff malice factor, $\tilde{\mu}$. Where the malice factor is below this cutoff, we have a standard revealing equilibrium, where weak types (here, $L$) always settle; and strong types ($H$) are taken to trial with a certain probability to maintain equilibrium. However, the existence of malicious plaintiffs, who have a greater appetite for trial, forces the defendant to take high offers to trial more often. The non-malicious type reluctantly engages in this equilibrium, and is bound to go to trial more often due to the inability to distinguish between malicious and non-malicious motivations.

The more interesting phenomenon occurs where the malice factor is sufficiently high. Then, the non-malicious type is induced to "reverse-mimic," and pool with the weaker types. This semi-separating result stands in contrast to the standard predictions. Moreover, as they pool together, these types can raise their settlement offer. This suggests that the weaker types, both malicious and non-malicious, are better off due to the spite motives; whereas the strong, non-malicious type is still worse off relative to the ordinary case.

## 8.2. Semi-Pooling Equilibrium: Characterization and Proof

**Proposition 3.** *Suppose $\bar{J} > \hat{J}$ and that $\dot{J} < \hat{J}$. Then, a perfect Bayesian equilibrium exists such that:*

$$(i) S^*(J) = \begin{cases} = T & \text{if } M = 0 \text{ and } J < \dot{J} \\ = \dot{S} = \dot{J} + c_D & \text{if } M = 0 \text{ and } \dot{J} < J < min\{J''_{NM}, \bar{J}\} \text{ or } M = 1 \text{ and } J < J''_M \\ > \bar{S} & \text{elsewhere.} \end{cases}$$

$$(ii)\ R^*(S) \longrightarrow \{Accept, Reject\} \begin{cases} (\dot{p}, 1 - \dot{p}) & \text{for } S = \dot{S} \\ (1, 0) & \text{for } S \leq T \\ (0, 1) & \text{elsewhere.} \end{cases}$$

$$(iii)\ B^*(S) = \begin{cases} (1 - \beta) \int_{c_P}^{\dot{J}} JdF & \text{if } S = T \\ \beta \int_{c_P}^{J''_M} JdF + (1 - \beta) \int_{\dot{J}}^{J''_{NM}} JdF & \text{if } S = \dot{S} \\ \beta \int_{min\{J''_M, \bar{J}\}}^{\bar{J}} JdF + (1 - \beta) \int_{min\{J''_{NM}, \bar{J}\}}^{\bar{J}} JdF & \text{elsewhere.} \end{cases}$$

*Where:* $\dot{p} = 1 - \frac{\dot{J} - c_P}{T}$ $\quad J''_{NM} = \dot{J} + T$; $\quad J''_M = \dot{J} + \frac{T}{1+\mu}$; $\quad \dot{J} = \frac{\beta \int_{c_P}^{min\{J''_M, \bar{J}\}} JdF + (1-\beta) \int_{\dot{J}}^{min\{J''_{NM}, \bar{J}\}} JdF}{\beta F(min\{J''_M, \bar{J}\}) + (1-\beta)(F(min\{J''_{NM}, \bar{J}\}) - F(\dot{J}))}$; $\quad \hat{J} = c_P + T\frac{\mu}{1+\mu}$

*Proof. i)* Consider non-malicious plaintiffs. Given the defendant's strategies, they face three options – reverse-mimicking to $T$ and settle with certainty; demanding $\dot{S}$ and being accepted with probability $\dot{p}$; and demanding something higher and getting rejected. Those with $J < \dot{J}$ find that their best option is to reverse-mimic to $T$, given that the payoff from demanding $\dot{S}$ is increasing in $J$ and that $\dot{J}$ itself is indifferent between the two options of demanding $\dot{S}$ and reverse-mimicking to $T$ (cf. Equation 7). Non-malicious plaintiffs between $\dot{J}$ and $J''_{NM}$ find that their best option is to reverse mimic to $\dot{S}$. If there are types above $J''_{NM}$, they prefer going to trial as their expected damages are very high. Accordingly, offering $S > \bar{S}$, which will be rejected, is a best option.

*ii)* Consider malicious plaintiffs, who have the same three options given the defendant's strategies. We have shown, assuming $\hat{J} > \dot{J}$, that all malicious plaintiffs prefer to mimic to $\dot{J}$ over reverse-mimicking to $S = T$ (and settling with certainty) (footnote 3 and accompanying text). However, as we have shown, very strong types, above $J''_M$, prefer certain trial and thus do not want their demands to be accepted, for instance, by demanding $S > \bar{S}$.

*iii)* Consider the defendant. If the defendant receives a demand of $T$ (or below), she has an incentive to accept this for certain ($T$ is the lowest liability she faces at trial). If she receives a demand of $\dot{S}$ she is, by Equation 9, indifferent between accepting and rejecting given her beliefs. Thus, accepting offers $\dot{S}$ at $\dot{p}$ is a best-response. Similarly, if she receives a proposal to settle for $\bar{S}$ she is better of rejecting it, as her liability from trial will (weakly) lower.

∎

## References

Abramowicz, M., 2024. Modeling fee shifting with computational game theory. GWU Legal Studies Research Paper No. 2024-35, URL https://papers.ssrn.com/sol3/papers.cfm?abstractid=4736839.

Bebchuk, L.A., 1984. Litigation and settlement under imperfect information. RAND Journal of Economics 15, 404–415.

Dari-Mattiacci, G., Saraceno, M., 2020. Fee shifting and accuracy in adjudication. International Review of Law and Economics 63, 1–15.

Daughety, A.F., Reinganum, J.F., 1994. Settlement negotiations with two-sided asymmetric information: Model duality, information distribution, and efficiency. International Review of Law and Economics 14, 283–98.

Deck, C.A., Pecorino, P., Solomon, M., 2024. Litigation with negative expected value suits: An experimental analysis. Journal of Empirical Legal Studies 21, 244–278.

Engel, C., Weinshall, K., 2020. Manna from heaven for judges: Judges' reaction to a quasi-random reduction in caseload. Journal of Empirical Legal Studies 17, 722–51.

Farmer, A., Pecorino, P., 2007. Negative expected value suits in a signaling model. Southern Economic Journal 74, 434–447.

Guerra, A., Lavie, S., Turrini, E., 2025. Information-signaling in pre-trial bargaining: An experiment, URL https://papers.ssrn.com/sol3/papers.cfm?abstractid=5219401, unpublished working paper.

Guha, B., 2016. Malicious litigation. International Review of Law and Economics 47, 24–32.

Guha, B., 2019. Malice in pretrial negotiations. International Review of Law and Economics 58, 25–33.

Klerman, D., Lee, Y.H.A., 2014. Inferences from litigated cases. Journal of Legal Studies 43, 209–48.

Lavie, S., Tabbach, A., 2023. Judgment-contingent penalties: Signaling in negative expected-value suits. The Journal of Legal Studies 52, 193–239.

Lavie, S., Tabbach, A.D., 2020. Judgment-Contingent Settlements. The Journal of Law, Economics, and Organization 36 (1), 170–206.

Pecorino, P., Van Boening, M., 2018. An empirical analysis of the signaling and screening models of litigation. American Law and Economics Review 20 (1), 214–244.

Reinganum, J.F., 1988. Plea bargaining and prosecutorial discretion. American Economic Review 78, 713–728.

Reinganum, J.F., Wilde, L.L., 1986. Settlement, litigation, and the allocation of litigation costs. RAND Journal of Economics 17, 557–566.

Schweizer, U., 1989. Litigation and settlement under two sided incomplete information. Review of Economic Studies 56, 163–77.

Shavell, S., 2004. Foundations of Economic Analysis of Law, Harvard University Press.

Spier, K., 2007. Handbook of Law and Economics (Vol. 1), Elsevier, chap. Litigation, pp. 259–342.