INTERNATIONAL COMPUTATIONAL LAW JOURNAL

CODEX SOURCE ARCHIVE

Will Robot Judges Change Litigation and Settlement Outcomes?

Anthony J. Casey, Anthony Niblett

Computational Law Report · Release 2.0 · 2020-08-14

Reader edition prepared by The Research University under CC BY 4.0. This cover was added; the original PDF pages follow unchanged. Original authorship, publication identity and source links are preserved. This independent republication is not sponsored or endorsed by Stanford University, Stanford Law School, CodeX or MIT.

Volume 1 · Issue 2 · Systems, Rules and Public Infrastructure

Original source and license evidence are recorded on the article landing page.

MIT Computational Law Report

Will Robot Judges Change Litigation and Settlement Outcomes?

Anthony J. Casey, Anthony Niblett

Published on: Aug 14, 2020

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

A First Look at the Algorithmic Replication of Prior Cases

0. Introduction

The promise (or threat) of so-called "Robot Judges" has captured the attention of popular media and legal scholarship.[1] Recently, the Estonian government made a splash by announcing its plan to use artificial intelligence to decide small claims cases.[2] And similar programs exist or are under development in China,[3] the Netherlands,[4] and other jurisdictions.[5] Despite the hype and attention, little has been said about what Robot Judges are or how they actually work.

Consider three potential candidates for automating judicial decisions: (1) structured code that explicitly stipulates "if x, then y," identifying whether certain factual elements are present, outputting a binary judicial decision; (2) advanced artificial intelligence that can determine the "right" outcome on its own; or (3) predictive litigation algorithms that use historical data from prior case outcomes to determine how the new case fits within the contours of existing law. The first method is of limited use —it can only be used to decide the very simplest of cases where there exists no ambiguity about the relevant factual elements. The second method requires technology that does not exist today. The third method, however, has promise. Litigation assessment algorithms do currently exist, and their prediction outputs could—at least in theory—be converted into automated judicial decisions. Still, the question remains: how does one go about converting predictive outputs into judicial decisions? This question, which has received little attention, is the focus of this paper. And, as it turns out, the answer has major effects on settlement and litigation outcomes.

This article, therefore, explores the effect of using litigation assessment algorithms in judicial decision making. These algorithms use data from past judgments to predict the likely outcome of a new case.[6] This technology is already used by lawyers,[7] and several scholars have pointed out its potential value as a judicial aid.[8] Indeed, some jurisdictions appear to already be using algorithms to guide judges.[9] And, in the extreme, the predictions could even be converted into automated judicial decisions.[10]

Here, we show that the use of algorithms to assist or automate judicial decision making can distort litigation and settlement outcomes.[11] The particular nature of the distortion depends on the methods judges actually use to translate predictions into judgments. Yet, while scholars have identified the advent of algorithm-assisted judging, few have explored the mechanisms by which a judge might translate the algorithmic output into an actual legal decision. We address that question. We use a simple model of litigation to explore various methods for converting algorithmic assessments into judicial decisions, and then demonstrate how the choice of method affects both the legal outcome of the case as well as the settlement dynamics among the parties.

Importantly for our analysis, legal decisions concerning liability are typically binary in nature—negligent or not negligent; guilty or not guilty. However, algorithmic assessments of liability are typically provided in probabilistic terms (for example, a 70% likelihood). On the surface, this distinction may seem unimportant, and converting probabilities into binary outcomes may *appear* to be a fairly intuitive exercise. After all, in other contexts, courts simply use a balance-of-probabilities or preponderance-of-the-evidence standard. But the conversion of probabilistic outcome predictions into binary legal decisions actually involves difficult choices with major consequences. And yet, researchers who study the interplay between algorithms and law – including ourselves – have hitherto overlooked these critically important effects and questions.

This article proceeds in four sections. The first section presents and defines the concept of litigation prediction tools. The second section presents a simple background model of litigation and settlement that will serve as a baseline comparison for our primary analysis. The third section presents various methods by which judges can use predictive tools and demonstrates how their choice affects litigation and settlement outcomes. The fourth section presents a general discussion.

1. Algorithmic Assessments of Liability

The possibility of courts using predictive algorithms to guide or even automate judicial decisions is real. And discussions around automated judging have garnered much attention in popular media, where the label of "robot judges" has caught on. Estonia has announced plans to implement a system of automated judgments for small claims, and the governments of other countries such as the Netherlands, the United Kingdom, and China, are reportedly implementing or looking into implementing similar systems. There are, however, unanswered questions regarding what it means for a decision to be rendered or even assisted by predictive algorithms or automated data analytics. Furthermore, the answers to those questions will affect outcomes and settlement dynamics in litigated cases.

In this article, we focus on the judicial use of one specific category of data-driven legal tools: algorithms that use data about past case outcomes to predict the outcome of a new or future case. We call any type of tool that fits this description an "algorithmic assessment of liability." These tools can provide an "objective" answer to the question: how does the existing law apply to the facts of the case at hand?[12]

For example, suppose a worker wishes to know whether she is an independent contractor or an employee for the purposes of employment law. In this instance, suppose that *if* the worker was found to be an employee, then she would be entitled to an additional payment upon leaving the firm. The legal question is a binary classification problem. In Canada alone, there have been over one thousand legal cases that have considered this question. Suppose now that the information in these cases can be

represented in the form of structured data. That is, every single case that has previously answered this legal issue has been coded in a structured way such that the existing law is represented in a dataset. Predictive techniques from statistics or machine learning can then generate the likely outcome — "*independent contractor*" or "*employee*" — for any new hypothetical set of facts. This is a simple binary classification algorithm. It generates a prediction of the most likely outcome as well as a probability associated with that outcome. That is: "In your case, the algorithm predicts you are an *employee* with an outcome probability of 70%."

At their core, these algorithmic assessments of liability use the data from past decisions to make predictions about how a human judge will apply the law to a given set of facts in the next case. But technology can take things a step further. The predictions might themselves be used by judges to inform their judgments in new cases. And in the extreme, the predictions could, themselves, be converted into an automated judicial decision. Others have recognized the possibility of this momentous step, but few have considered how it would work. As we demonstrate below, this conversion creates complicated choices with major effects on case outcomes and thereby affects the *ex ante* behavior the law is intended to regulate. Indeed, even if one assumes an unrealistically simple environment with a perfectly accurate algorithm, the usage of algorithmic assessments to automate or merely inform judicial decisions can dramatically change litigation and settlement outcomes.[13]

While we assume accuracy in our model, we do not claim that such accuracy exists in the real world. It is well understood that logistical challenges make accurate predictions difficult to achieve. For example, these algorithms require data from a sufficiently large number of cases dealing with the relevant factual situations and presenting a sufficient level of consistency. A more heterogeneous area of law will result in lower out-of-sample prediction accuracy. A low number of cases will result in uncertainty about the prediction. And the usual issues of selection bias, publication bias, and coder bias present obstacles to achieving a completely accurate series of predictions.

These questions of accuracy and bias are important. For the purposes of this paper, however, we bracket them. Our goal here is to highlight another important challenge that exists *even if* the problems of accuracy and bias were solved. Assuming that the predictions are "accurate" in their probabilistic assessment of liability allows us to isolate the challenge of translating the algorithmic outputs into judgments, which has been ignored in the literature. As with accuracy and bias, this problem is worthy of attention. Despite objections based on accuracy and bias, litigation prediction tools are currently available and are being used by law firms, insurance companies, and litigation finance firms in many contexts. The move to use these technologies in some courts is already underway, and will only grow as more data becomes available, technologies improve, and accuracy increases. It is important, therefore, to understand the issues concerning translating predictions into

judgments before judicial use of algorithms becomes widespread. This paper is a first step toward gaining that understanding.

2. A Baseline Model of Settlement and Litigation

To see how algorithmic assessments of liability change litigation and settlement outcomes, it is first necessary to establish the baseline—how cases are currently decided and settled in the absence of algorithmic assessments. We begin therefore with a simplified baseline model, derived from the existing literature where settlement occurs unless there is asymmetric information or optimism on the part of at least one of the parties to the litigation.[14]

For simplicity, we focus on questions of law.[15] We assume that the parties have agreed upon (or at least stipulated as to) a set of facts and are asking a judge to apply uncertain law to those facts. In this way, the question can be viewed as akin to a conventional motion for summary judgment or an appeal of a legal issue.[16] Because the facts are stipulated, the only information the parties do not already know is the court's interpretation of the law.

Models of settlement and litigation in the literature of law and economics typically begin with the assumption of a probability of plaintiff success.[17] We include in our model the idea of an "accurate" probability of plaintiff success, p^I. This represents the most accurate representation of the objective probability of plaintiff success given the stipulated set of facts. Later, p^I will become relevant when we assume that the predictive algorithm can reveal the probability to the parties.

The baseline model in this section begins in a world where no party or legal decision maker has access to algorithmic prediction tools. A legal battle in our model is fought between two risk-neutral parties, P and D. P is the plaintiff. She claims damages of L from D, the defendant. The magnitude of L is not in dispute. The dispute lies only in the question of legal liability: given the stipulated facts, is D liable for the losses of P?

P, the plaintiff, believes she has probability p^P of winning the case. D, the defendant, believes that P has a probability p^D of winning. These are subjective probabilities. We also assume that the parties' subjective probabilities are, to different extents, correlated with the accurate and objective probability, p^I. This assumption captures the real-world fact that the parties each possess valuable but imperfect information about the case. With that information they are unable to know p^I exactly, but their estimates of p^P and p^D will reflect some valuable information tying them loosely to p^I . We also assume that both parties are optimistic relative to the objective probability of plaintiff success. To keep things simple, we therefore assume the following relationships:

$$p^P = p^I + \delta$$

$$p^D = p^{\mathrm{I}} - \epsilon$$

This assumption has little importance within the baseline model, but it will become important later on.[18]

We define Δp as the area of disagreement: the difference between the two subjective probabilities of the parties, $p^P - p^D$ (which, by definition, equals $\delta + \epsilon$).

We presume the costs of reaching a settlement are zero, but the costs of a legal adjudication are positive. The costs for each party are c^P and c^D. For simplicity, these costs are not endogenous. The costs are fixed and are only incurred if settlement fails and the parties go to court. We also assume the "American rule" for awarding cost, where each side bears their own costs.[19] Further, we suppose that each party has information about the other party's beliefs and costs.

The plaintiff and the defendant play a simple game.[20] The timing of this game is as follows:

- Stage 1 – *P* decides whether or not to bring suit.
- Stage 2 – *D* makes a "take-it-or-leave-it" offer of settlement, S^D.
- Stage 3 – *P* decides to accept or reject *D*'s offer. If she accepts, the game is over. If she rejects, both parties pay their respective costs, c^P and c^D, and liability is determined by a neutral third-party judge, *J*. The judge is not a player in the game, but merely determines liability after applying the law to the facts. If the judge finds in favor of *P*, she awards compensatory damages of *L*, which must be paid by *D*. In line with our assumption that p^I is an accurate assessment of the likelihood of liability, *D* is found liable with probability p^I.

The equilibrium of this game is simple and intuitive. Working backwards, in stage 3, *P* will accept *D*'s offer if and only if the settlement offer *S* is greater than the net return *P* would subjectively expect to receive from going to court. That is, *P* will settle if and only if:

$$S^P \geq \mathrm{p}^P L - c^P$$

The defendant, in stage 2, will offer a settlement amount up to her net expected losses from litigating the case. The defendant will never offer more than the losses that she subjectively would expect to pay at court:

$$S^D \leq \mathrm{p}^D L + c^D$$

We frame the equilibrium in terms of the difference in subjective probabilities. If the two parties have relatively similar assessments of the merits of the plaintiff's case, then it will be in the interest of both parties to settle. If, however, the parties substantially disagree – and p^P is much greater than p^D – settlement will not take place and the parties will proceed to court.

Settlement only occurs if this condition, equation (1), is satisfied:

$$\Delta p \leq \frac{c^P + c^D}{L}$$

For the sake of simplicity and consistency throughout the rest of the paper, we call the right hand side of this equation the "cost-damage ratio." **Figure 1** illustrates the case where the area of disagreement is too large for settlement to be possible at equilibrium.

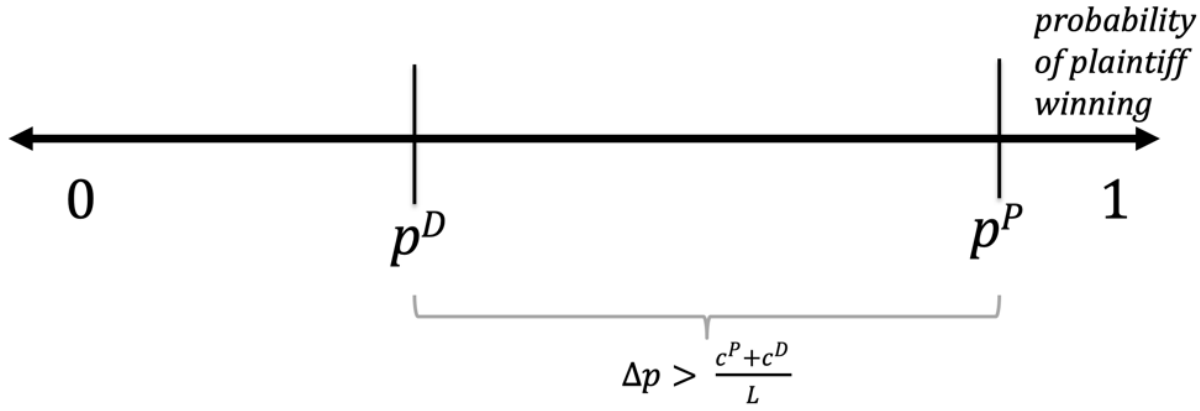


Figure 1: The subjective beliefs of the plaintiff and defendant are separated by an area of disagreement greater than the cost-damage ratio. Under these conditions, the two parties are unable to reach a settlement.

But what about when the area of disagreement is not greater than the cost-damage ratio? That is, what actually happens when parties settle? The dynamics of settlement in this model are very straightforward. If the settlement condition in equation (1) is satisfied, the defendant in stage 2 makes a take-it-or-leave-it offer, S^*, where $S^* = p^P L - c^P$. At equilibrium, this offer is accepted by the plaintiff in stage 3. The dispute is settled; it does not proceed to a judge.

Given our assumption about the correlation between the subjective probabilities p^P and p^D and the objective probability p^I, the magnitude of the settlement offer is also correlated to the objective probability of the plaintiff's success:

$$S^* = (p^I + \delta)L - c^P$$

This relationship between S^* and p^I is continuous. The greater the likelihood of the plaintiff winning, the larger the settlement amount. This is illustrated in **Figure 2**.

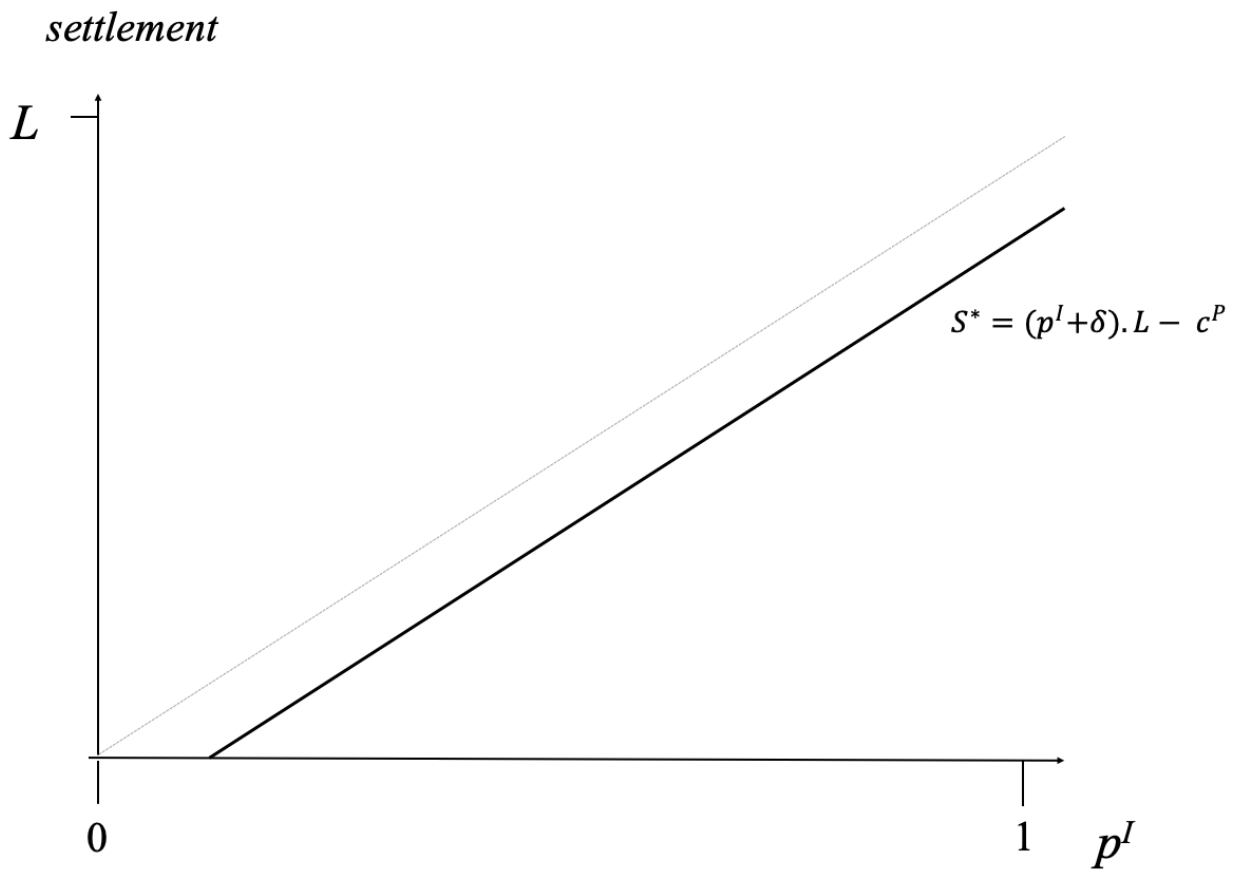


Figure 2: Settlement offers are correlated with p^I. The relationship is linear. Offers will be accepted if the condition in equation (1) is satisfied.

We will refer to this as "the baseline model with subjective probabilities." This information is not new. These results can be found, in one form or another, in the existing literature on law and economics regarding litigation and settlement.[21] We wish, however, to emphasize two key takeaways. First, settlement fails within this model when the plaintiff's subjective view of winning, p^P, is sufficiently greater than the defendant's belief that the plaintiff will win, p^D. If both parties are overly optimistic, then no settlement offer is accepted at equilibrium. Second, when settlement does occur, the magnitude of the settlement offer will be a function of the objective probability of liability, p^I. That is, $S^* = f(p^I)$, where $f'\left(p^I\right) > 0$. The greater the objective probability liability, the larger the settlement offer.

From here it is uncomplicated to demonstrate what results when the parties themselves have access to predictive algorithms that reveal p^I. If the parties understand that the algorithm is revealing the accurate objective probability, and they both believe the objective probability, then every case settles with a settlement payment tied to p^I:

$$S_i^* = p^I L - c^P$$

We will refer to this as the "baseline model with objective probabilities." This is illustrated in **Figure 3**.

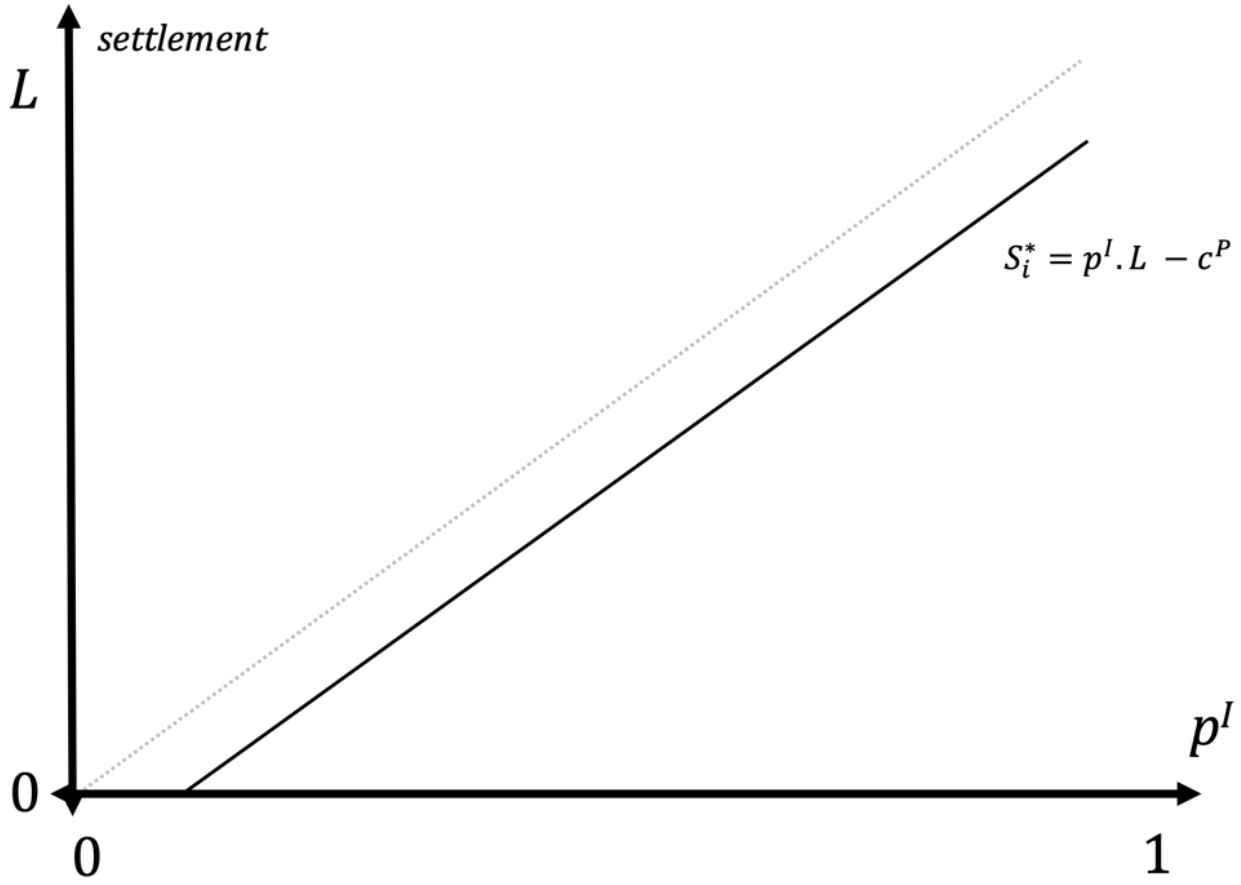


***Figure 3*:** The baseline model with objective probabilities

More realistically, the parties will have some doubts about the algorithm and thus assume that the output is a valuable but imperfect estimate of p^I. In that case, the parties will update their subjective priors, p^P and p^D, based on the weight they place on the algorithmic output. In most cases, this will increase the likelihood of settlement. Though in some "atypical" cases, it may reduce settlement outcomes. These results are not essential to our analysis, but we have provided them in the Appendix.

Less straightforward is what occurs when a judge has access to predictive algorithms that reveal p^I. In the next section, we introduce this scenario and show how it affects litigation and settlement outcomes.

3. Judicial Use of Algorithmic Assessments of Liability

We now introduce algorithmic assessments of liability into the model. We suppose that the algorithmic assessment of liability is accurate and provides the judge with an independent and objective probability of a plaintiff succeeding, p^I. But the probabilistic assessment is not a legal decision. It becomes an input into a legal decision. When a judge takes notice of p^I, she may translate the

probability into a final judgment in various different methods. The choice of method has major implications. It can change the litigated outcomes and alter fundamental characteristics of an adjudication system. Moreover, it can alter settlement dynamics.

Before we discuss the different options available to judges to convert these algorithmic predictions into liability rules, we wish to point out some of the tradeoffs of using such tools to make decisions. First, if judges use algorithmic assessments of liability as the basis for legal decisions, this implies that judges are satisfied that the previous decisions are accurate representations of what is "right." That is, judges must be comfortable with the content of existing law if they are prepared to use these types of algorithms. If, for some reason, judges believe that the existing law should be changed, reliance on these algorithms would not be sensible.

Second, we have made the rather strong assumption that the algorithms are "accurate." A prominent reason they may be inaccurate is that datasets that describe and explain the law are often based on existing case law, which are themselves subject to selection bias. Not all disputes turn into legal disputes; not all legal disputes are litigated; not all litigated cases are determined by judges; and not all judges' decisions are published.

With those caveats, we present six basic options for converting probabilistic assessments into legal outcomes. Other options may be available, but this analysis provides an initial examination of the most likely options to be proposed. For each option, we show how the use of algorithmic assessments of liability changes both litigation outcomes and settlement dynamics. We assume for each option that the litigants also have access to the prediction tools.[22]

3.1 Option 1: Binary Outcome Determined by *Ex Ante* Balance of Probabilities

The first instinctive reaction many people have when talking about automated judgments is that defendant liability will be based upon whether the data tells us the defendant is *more likely than not* liable.[23] That is, if the prediction tool says it is 51% likely that the plaintiff has established liability, then the plaintiff wins, and full damages are awarded. If the prediction tool says it is only 49% likely, then the defendant wins, and no damages are awarded. Thus:

- If $p^I \leq 50\%$, the defendant wins, and the plaintiff is awarded zero damages;
- If $p^I > 50\%$, the plaintiff wins and is awarded L.

These outcomes are represented visually in **Figure 4**.

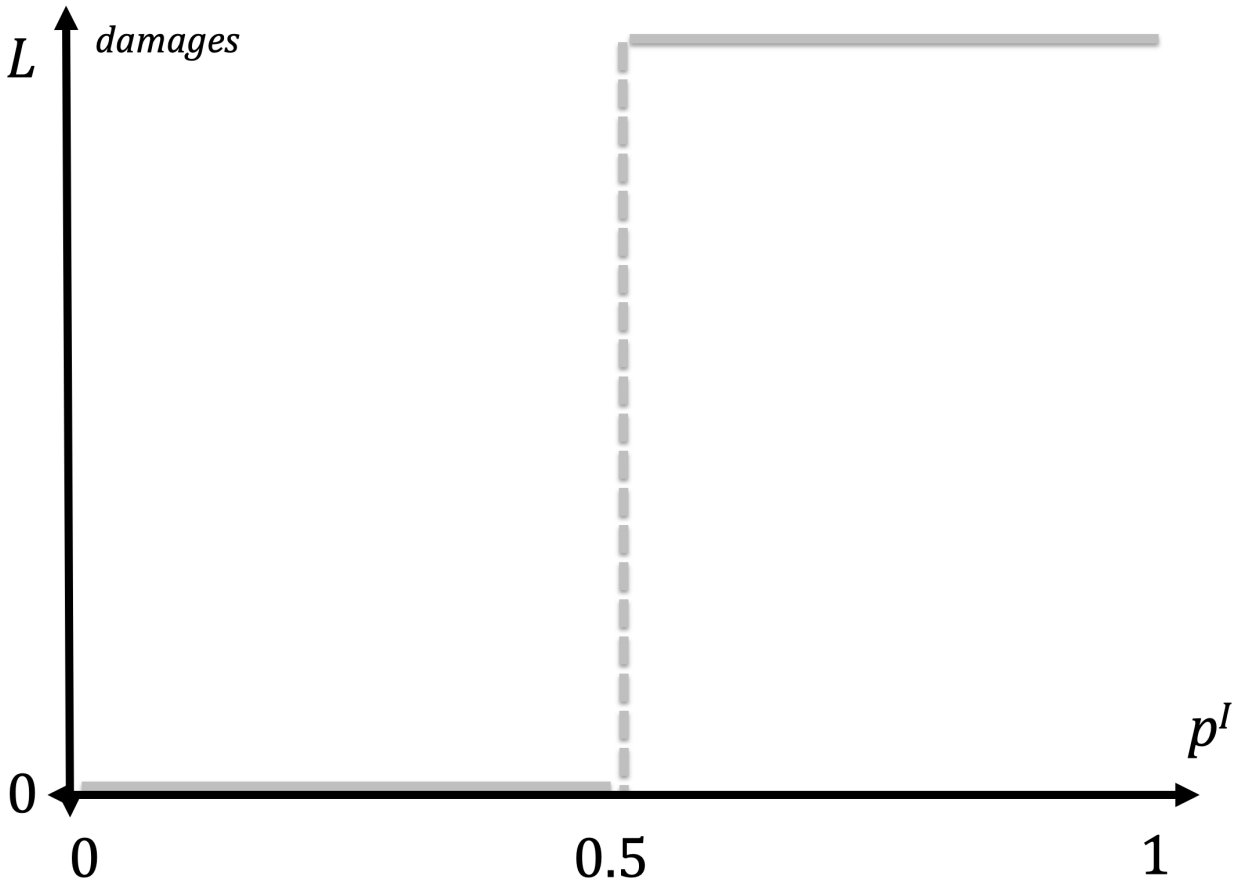


Figure 4: Where p^I is less than or equal to 50%, the defendant wins. No damages are awarded. Where p^I is greater than 50%, the plaintiff wins and is awarded full damages, *L*.

The intuition behind this mechanism may be its similarity to burdens of proof in the context of fact finding. But those concepts do not fit here. The *more likely than not* standard that lawyers know so well specifically concern factual determinations. But we are asking for a prediction about legal determinations. Judges do not announce the law in terms of a likelihood standard.[24]

3.1.1 Litigation Outcomes

How does the use of algorithmic assessments compare to litigation outcomes without the algorithm? When a judge decides a case without the assistance of an algorithm, a plaintiff's expected return is $p^I \cdot L - c^P$. (Recall that, by assumption, p^I is both the output from the algorithm and the accurate objective probability).

The outcomes differ when judges use the binary model of algorithmic assessment. The outcomes abruptly change as the liability probability nears 50%. This transforms the outcomes of cases. Take, for example, a case where the plaintiff is assessed to be 70% likely to be found liable (p^I = 70%). Remember that without the algorithm, when cases go to court and are decided by human judges, 30% of them result in a defendant victory. But, under the binary model of algorithmic assessment, 0% of

cases result in defendant victory. Thus, 30% of cases are decided differently under the binary model. As a further example, if p^I were equal to 10%, then the binary model of algorithmic assessment gives a different outcome in 10% of cases.

The quantity of cases decided differently is greatest where the algorithm predicts that the case is "close"—when p^I is close to 50%. The exact proportion of all cases that are "decided differently" depends on the underlying distribution of cases under p^I. But, for sake of explanatory ease, suppose that the distribution of cases that go to court follow a uniform distribution from $p^I \in [0, 1]$. Under a uniform distribution, the final outcome regarding liability would be different in one-quarter of all cases.[25]

3.1.2 *Settlement Outcomes*

Mapping probabilities into binary outcomes in this way radically changes the nature of settlement. Settlement outcomes track expected outcomes from the predicted probabilities. Using backward induction, consider how the plaintiff behaves in stage 3. If $p^I \leq 50\%$, then the plaintiff knows that she will pay c^P to go to court and has no chance of winning. Thus, in stage 3 she is willing to settle for any amount greater than zero. If $p^I > 50\%$, the plaintiff knows that she will win and recover $L - c^P$ if she rejects settlement. Knowing this, in stage 2, the defendant will offer zero if $p^I < 50\%$, or $L - c^P$ if $p^I > 50\%$. The plaintiff will not bring a claim in stage 1 if $p^I < 50\%$. The equilibrium can be stated simply:

1. No cases are brought by the plaintiff when $p^I \leq 50\%$
2. All cases where $p^I > 50\%$, the plaintiff settles the case for $S = L - c^P$

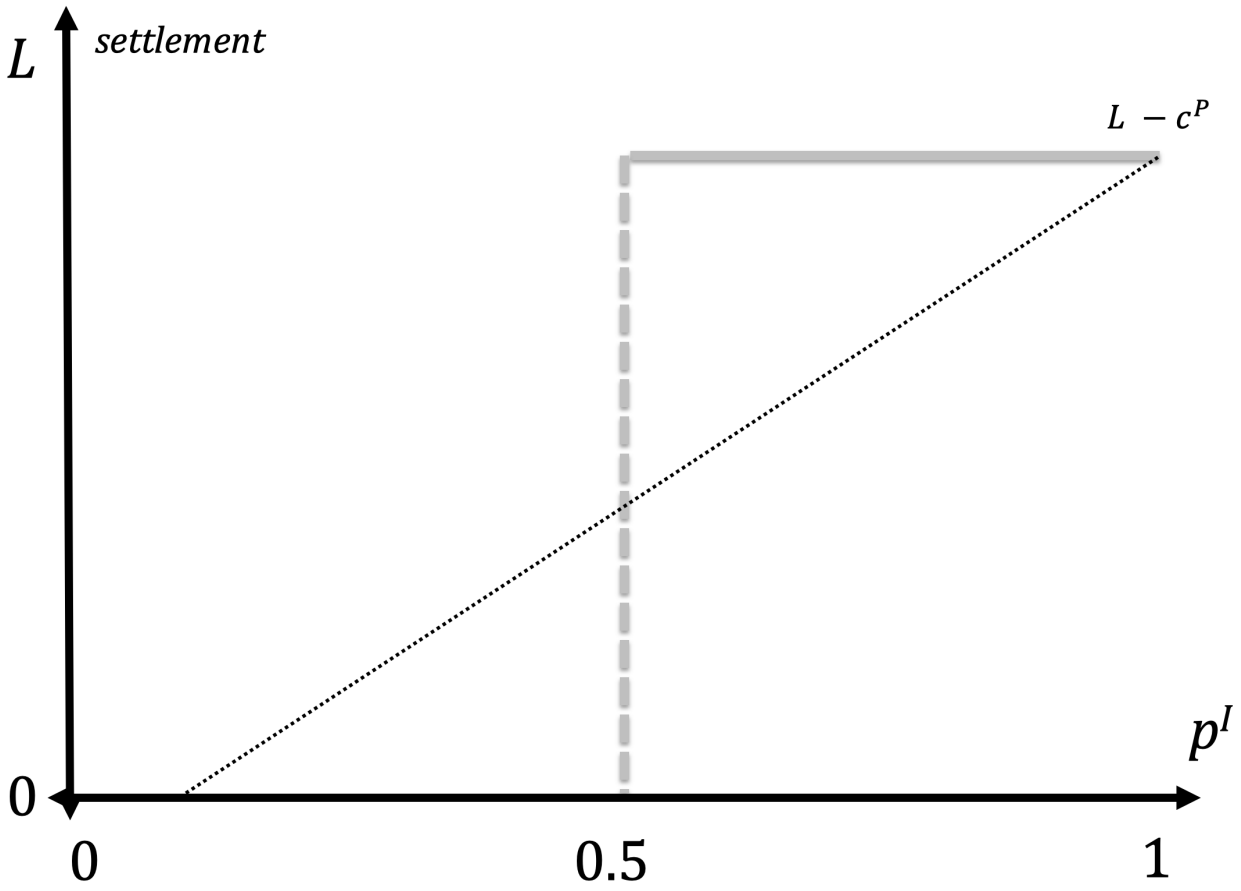


Figure 5: Where p^I is less than or equal to 50%, the case is not brought. Where p^I is greater than 50%, the case is brought and the plaintiff settles for the full amount minus costs.

Accepted settlement offers, in equilibrium, are shown graphically in **Figure 5**. The thick gray line represents the amount the plaintiff recovers in a litigation environment where judges use algorithmic assessment to determine liability. Where $p^I \leq 50\%$, the plaintiff does not bring a case. At $p^I = 50\%$, there is a vertical spike. For all probabilities above 50%, the plaintiff recovers (nearly) the full settlement amount. Compare the thick gray line to the dotted black line. The dotted black line might represent the settlement outcomes in the baseline model using subjective probabilities (see **Figure 2**) or the baseline model using objective probabilities (see **Figure 3**). There are four important differences between the outcomes here and the outcomes in the baseline models:

1. In both baseline models, the dotted black line "tracks" the independent assessment of the probability of the plaintiff being liable. The gray line also tracks the probability, but in much coarser fashion. Before the binary model of algorithmic assessment was used, the settlement behavior of litigants did not distinguish greatly between a 49% plaintiff and a 51% plaintiff. But in a world where the shift at 50% becomes law, the difference between these two cases is stark. At 51%, the plaintiff recovers as if she were at 100% within the baseline model with objective probabilities. At 49%, she recovers nothing. Further, in a situation where judges do not use algorithmic tools, a defendant who

is 100% likely to be liable behaves very differently in settlement to a defendant who is 51% likely to be liable. Under these circumstances, they have the same incentives to settle.

2. In the baseline model with subjective probabilities, without algorithmic aids, the dotted black line is *conditional* on settlement being possible. Settlement depends upon the area of disagreement between the two parties being sufficiently small. The gray line here comes with no such conditions. Settlement always occurs when a case is brought because there is no *ex ante* uncertainty about the outcome of liability.
3. In the baseline model with objective probabilities, parties settle every case only because all parties trust the algorithm to be accurate. Such trust is not required here. The parties' belief about the algorithm's accuracy is unimportant. The only requirements are that the parties know the court follows the algorithmic output and what that output is.
4. Plaintiffs do not bring disputes when the probability is less than 50%. Cases either settle near full liability or they are not brought at all.

3.1.3 Effect on Ex Ante Behavior

To the extent liability is a deterrent with regard to *ex ante* behavior, the switch from the baseline model to the binary model of algorithmic assessment is significant and problematic. Switching models essentially creates an inefficient liability cross-subsidy from defendants just over the 50% mark to those just below it. This will likely lead to a discontinuity in *ex ante* behavior where potential defendants tend to cluster their behavior around 49% and completely avoid behavior that is around 51%. Someone who is doing something at the 51% mark has strong incentives to alter their behavior toward the 49% mark.[26] On the other hand, someone who is doing something at the 49% mark has no expected liability, and therefore no incentive to expend any effort to move from 49% toward 1%. Similarly, someone who is at 100% percent liability has no incentive to expend any effort to move toward 51%. Thus, the binary model significantly modifies deterrence dynamics and cannot be viewed as merely an automation of the existing system. This is a rather dramatic shift in the substantive effect of the law.

3.2 Option 2: Continuous Outcomes – Expected Damages

An alternative proposal is to award expected damages. If the plaintiff rejects the defendant's settlement offer, the judge uses the independent assessment to award damages of $p^I \bullet L$ (shown graphically in **Figure 6**).

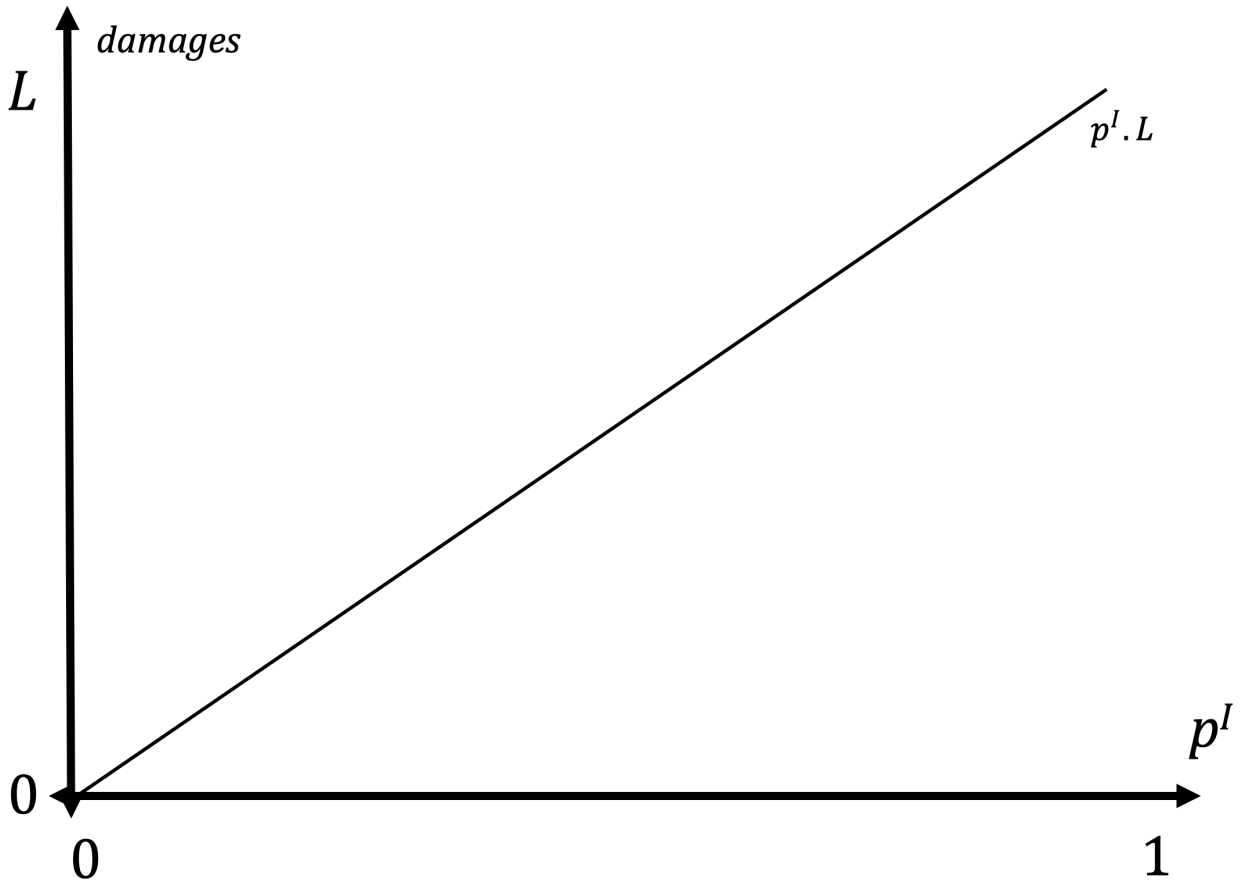


Figure 6: Damages awarded are expected damages, based on the independent assessment of the plaintiff's case.

3.2.1 Litigation Outcomes

This is an intuitive application of the data, but directly assigning outcomes from expected probabilities would represent a radical change in the way we think about liability and compensation in the Anglo-American legal system. Importantly, it holds defendants liable (partially) for losses even when there is a low likelihood of being found liable. If the algorithm stipulates that the plaintiff has a 5% chance of winning on the merits, should we hold the defendant liable for 5% of the losses? On the one hand, it may conflict with our intuitions to hold defendants liable when there is such a small chance of the plaintiff's case succeeding (even though they are only liable for a small percentage of the loss). On the other hand, as we shall see, this option simply entrenches into law what already happens with settlement in circumstances where parties (but not judges) use the algorithm and execute accurate Bayesian inferences.

3.2.2 Settlement Outcomes

Here, the settlement equilibrium is straightforward. The defendant offers $p^I \bullet L - c^P$ in stage 2, which is always accepted by the plaintiff in stage 3. (The plaintiff only brings cases where $p^I \bullet L - c^P > 0$.) This settlement offer by the defendant is the same as the offer in the baseline model with objective

probabilities shown graphically in **Figure 3** above. At equilibrium, all cases settle. The settlement offer, at equilibrium, reflects the probability of the plaintiff winning. The plaintiff who has an 80% probability of winning recovers 80% of the loss (minus costs).

3.2.3 Effect on Ex Ante Behavior

Beyond that, this method does not change *ex ante* incentives or deterrence, and so the expected damages model can be viewed as merely adding information to and automating results from the existing system. In other words, it produces the same effect on *ex ante* behavior as the baseline model with objective probabilities.

3.3 Option 3: Hybrid Approach

A slightly modified approach could combine the first and second options. Here, the defendant would be liable for *expected* damages if $p^I > 50\%$ and no damages if $p^I \leq 50\%$, the defendant wins.

3.3.1 Litigation Outcomes

This approach is perhaps more consistent with Anglo-American traditions. It is graphically represented in **Figure 7**.

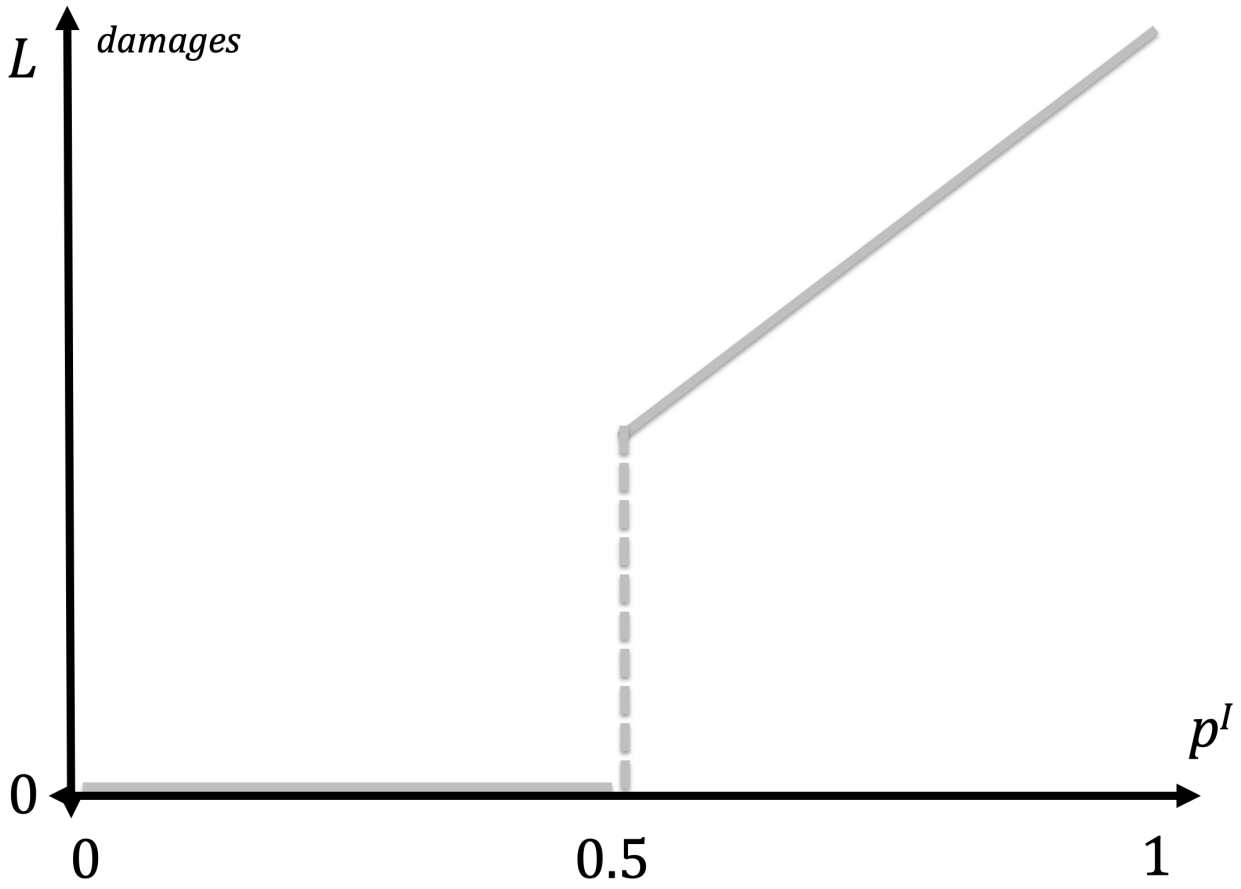


Figure 7: No liability is found when p^I is less than or equal to 50%. Where p^I is greater than 50%, expected damages are awarded.

3.3.2 Settlement Outcomes

The settlement outcomes are straightforward. They track the judgment outcomes, less costs.

3.3.3 Effect on Ex Ante Behavior

While this may seem like an acceptable compromise, converting predictions into outcomes would not produce adequate deterrence in *ex ante* behavior. Cutting off the possibility of liability for all cases where $p^I \leq 50\%$ reduces the *ex ante* expectation of compensation to be paid by the defendant. As a result, adopting this model amounts to a major change in the substantive effect of the law.

3.4 Option 4: Probabilistic Liability with Full Damages

As a fourth option, suppose that courts simulate a world without algorithms through probabilistic liability. Thus, when p^I is 70%, the court would impose full liability on the defendant 70% of the time. The idea is like flipping a weighted coin to determine liability, where the weight of the coin is determined by p^I. In the 70% example, the coin would be weighted land heads up (as in, assign liability) 70% of the time.

3.4.1 Litigation Outcomes

This approach would restore the expected liability outcomes that exist in the baseline model with objective probabilities. Graphically, the outcomes of judgments here look the same as in **Figure 6**.

Even though expected liability is the same as in the baseline model with objective probabilities, this approach would likely be met with heavy opposition. One might object that the rule of law is violated by allowing such random and arbitrary considerations to determine liability.[27] Assessing whether the approach represents an improvement requires consideration about the existing variation in judicial rulings. In practice, this option converts *unexplained* variation in legal decisions into *purely arbitrary* variation.[28] If one thinks that the unexplained variation is a result of nefarious bias, this is an improvement. If one thinks that the unexplained variation is a result of arbitrary judicial whims, nothing has changed—you have replaced one arbitrary method with another. But, if one thinks that behind judicial variation there are valuable case-specific human judicial intuitions that cannot be explained by data, then this would produce worse results by ignoring those justifications. In reality, the truth probably involves a combination of these factors. Any particular decision likely results from a mix of measurable information, biases, arbitrary distinctions, and unmeasurable intuitions.

3.4.2 Settlement Outcomes

Importantly, however, none of the critiques about outcome variation from the prior subsection matter if cases settle. And, in this model, all cases do settle. The parties have access to the algorithmic assessment and no disagreement about the outcome. As a result, cases will settle at the expected liability point (minus costs), thus creating the same settlement rates as option 2 and as in the baseline model with objective probabilities.

It is worth emphasizing that when settlement is ubiquitous, the source of variation that we discussed in the previous subsection is unimportant. Rational actors settling a case only care about the expected rate of liability.[29] As long as they are powerless to change that rate, they do not care about its causes.

3.4.3 Effect on Ex Ante Behavior

This result does not change *ex ante* incentives and deterrence. The probabilistic liability model produces the same effects on behavior as the expected damages model and the baseline model with objective probabilities.

3.5 Option 5: Triage “Easy” Cases

Another intuitive option for mapping probabilistic liability onto legal outcomes is to triage “easy” cases with predictive tools. That is, the judge can use the independent assessment, relying on algorithms to determine liability, only when the case is “easy.” Easy cases are those where the probability of one side winning is close to 100%. If the independent probability assessment is close to 0% or 100%, then the

outcome can be determined *ex ante*. In this way, courts can reduce their caseload by triaging easy cases from their list, focusing on those cases where the outcome is less clear.

Take, for example, a situation where the judiciary uses a 5% threshold: cases where one side has less than a 5% chance of winning according to the independent assessment are automatically determined by the independent assessment. If p^I < 5%, then the plaintiff recovers zero; if p^I > 95%, then the plaintiff recovers the full amount. For all other cases, where $p^I \in [5\%, 95\%]$, the case proceeds to human judicial determination if settlement fails. For the subset of easy, triaged cases, the litigation outcomes mirror those in option 1. For those cases where the algorithm is not used, the judge has leeway to decide the outcome. The larger the subset of triaged cases – that is, as the threshold for what is considered an "easy" case becomes more inclusive – the more this option begins to reflect the litigation outcomes of option 1. An example with a 5% threshold is shown graphically in **Figure 8**.

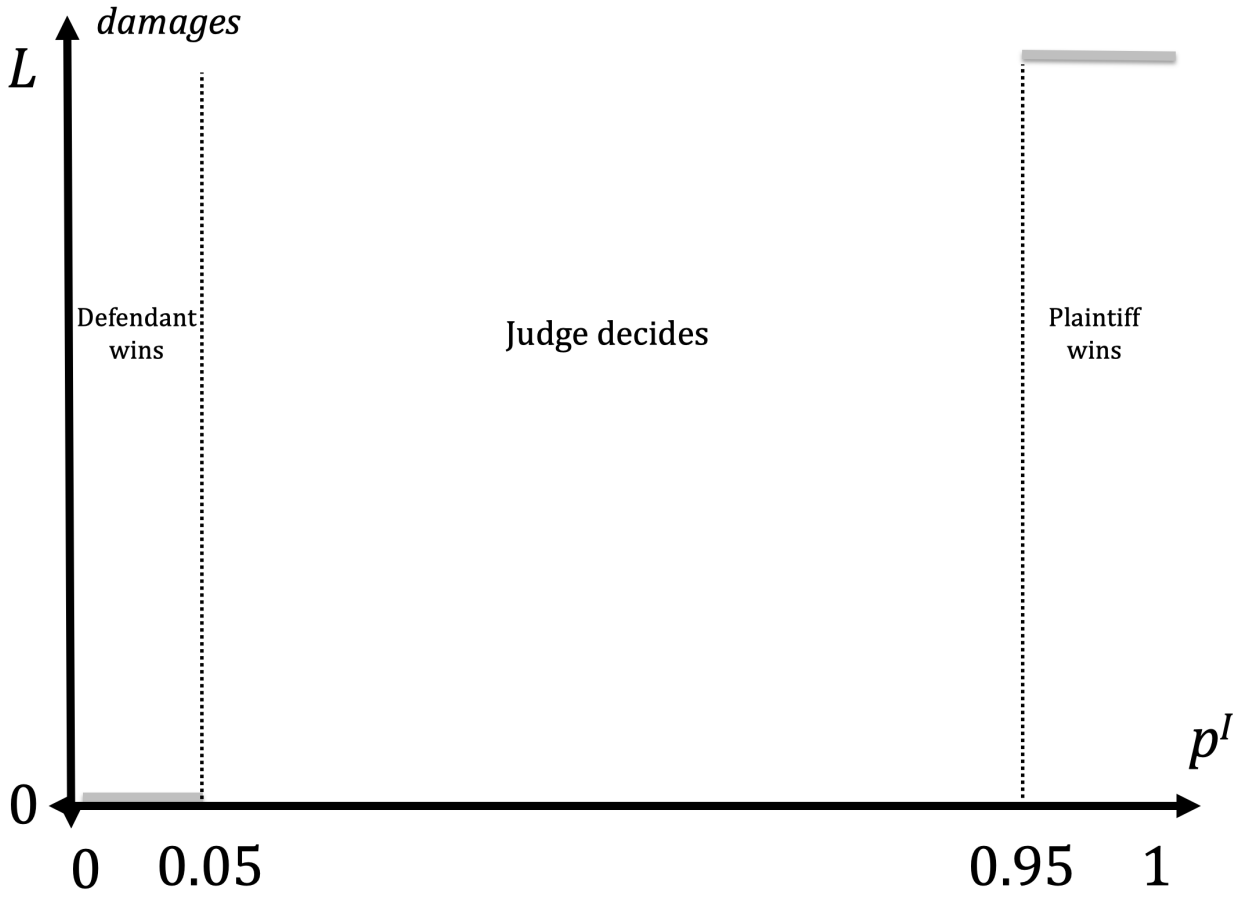


Figure 8: Easy cases, where one party has less than 5% of winning, are triaged. Liability in the intermediate cases is determined by the judge.

3.5.1 Litigation Outcomes

The number of cases that are "decided differently" in this option is clearly reduced compared to option 1.[30] Because cases at the two extremes are highly likely to be decided the same way, the distortion in

outcomes by using the algorithm is minimized. Of course, as the definition of "easy" expands, the greater the proportion of cases that are decided differently.

Many perceived benefits of triaging cases in this way may not be readily observable given our simple model. Frivolous suits (where the probability of a plaintiff winning is zero) are possible in the real world, even when the plaintiff knows the probability is zero.[31] Frivolous suits don't occur at equilibrium in our model, but it is not difficult to generate models to account for them. Algorithmic triage in these types of cases would effectively deter frivolous lawsuits.

3.5.2 Settlement Outcomes

The effect of triage on settlement depends on whether the independent assessment falls within the triaged zone. Let's say the threshold is 5%. If $p^I < 0.05$, the case is never brought. If $p^I > 0.95$, then the case settles for close to the full amount claimed. For intermediate cases, settlement results would follow our analysis in the baseline model with subjective probabilities.

3.5.3 Effect on Ex Ante Behavior

Similarly, the effect on *ex ante* behavior depends on whether a case is in the triaged zone. *Ex ante* behavior will only change if the suit is triaged. But this change might be small. Moving a 98% liability expectation to 100% and a 2% liability expectation to zero is a small – perhaps trivial – change in *ex ante* expectations. For the intermediate cases, the effects again remain the same as in the baseline model with subjective probabilities.

3.6 Option 6: Judges Use the Prediction for Guidance Only

The previous five options have all involved human decision makers deferring to the algorithmic assessment to some degree. For options 1 through 4, the algorithm is the ultimate arbiter of all cases, and for option 5, the algorithm is the arbiter of easy cases. But judges are unlikely to completely defer to the algorithm initially. Rather, at the outset judges will likely have much discretion to accept or reject the algorithmic assessment.

In option 6, we suppose that the judge has access to the algorithmic assessment of liability, but there is no requirement to rule based upon it. There is no formalistic mapping of probabilities onto outcomes. Instead, the judge has the option of referring to the probability assessment, but only uses the algorithm's assessment for guidance.

3.6.1 Litigation Outcomes

How will this change litigation outcomes in the event that cases go to court? The degree to which this affects outcomes will depend on the individual propensities of the judge using the tool. Suppose the judge ignores the recommendations of the algorithm? In that case, very little will change. The judge

relies on her own assessment of the case, as indeed she would in the absence of any such prediction. But, to the extent that the judge does begin to lean on the assessments of liability, litigation outcomes will be affected depending on which of the five previous options best describes how the judge is using the tool.

3.6.2 Settlement Outcomes

If parties know that the judge has access to the algorithmic assessment, but don't know her exact propensity to follow the guidance of the algorithm, how will this affect settlement? Upon revealing the algorithm's prediction, the plaintiff and defendant must reassess their prior assumptions knowing that the judge too has access to the prediction.

The tool provides an independent assessment of the likelihood of a plaintiff succeeding, p^I. Before stage 1, both parties update their prior subjective probabilities to reflect how they believe that the judge will update her position:

$$p_i^P = (1-\theta)p^P + \theta p^I$$
$$p_i^P = (1-\theta)p^P + \theta p^I$$

The degree to which parties update, θ or τ, depends on the degree to which they each believe the judge will rely on the algorithm. If parties believe that the judge will be highly influenced by the independent assessment, then θ and τ are close to 1. If parties are skeptical about the judge using the algorithm, θ and τ are close to 0.

When judges use the algorithmic assessments of liability, τ and θ represent the parties' beliefs about whether and how judges will update their determinations. In short, the plaintiff and defendant will adjust their subjective probabilities based on their beliefs about (1) how much the judge will update her ruling; and (2) how the judge will incorporate the probabilities into her decision. If parties believe that the judge will fully update her prior perspective and use the binary model (in 3.1), then there is almost a self-fulfilling prophecy – the settlement offers at equilibrium mirror the sharply disparate outcomes in option 1.

Depending on how parties perceive the judge's basis for decision, we may witness counterintuitive effects on settlement. For example, let's say that the plaintiff initially believes that she has a 90% chance of winning. However, the independent algorithm suggests that chance is only 60%. If the plaintiff believes that the judge will be faithful to the prediction *and* believes that the judge will employ the binary model, then the plaintiff will adjust their confidence level from 90% to a posture of 100%. If only the parties used the prediction tool, the equilibrium settlement offered by the defendant and accepted by the plaintiff would likely *decrease* upon revelation of a lower objective probability.[32]

But when the judge has access to the tool, the equilibrium settlement offer may *increase* upon revelation of a lower objective probability.

4. Discussion and Extensions

Settlement models in law and economics are based on the plaintiff's probability of victory.[33] The output of a litigation prediction tool follows these probability-based models as well. But what, exactly, does that probability mean? We might think of the probability of p^I = 80% in a tort case as meaning the following: If that case were litigated 100 times, the court would on average find in favor of the plaintiff in 80 of those cases. That statement does not conform to ideas like burden of proof for factfinders. It is not equivalent to saying there is an 80% chance the defendant committed the tort. To see why, imagine a case where everyone in the world agrees that factually there is a 60% chance the defendant committed the tort. Liability would be found in that case 100% of the time. In another case, imagine that half of the judges in the world think the defendant is 99% likely to have committed the tort and the other half think the defendant is 49% likely. There will be liability in 50% of the cases in this scenario.

Because these concepts are distinct, when technology attempts to translate the probability of a plaintiff's success into a judicial rule, mechanically importing a 51% threshold for liability is inappropriate. As we have shown, the precise mapping or translation from p^I into liability and damages can have important effects on litigation and settlement outcomes. Additionally, we have shown that even the mere use of prediction tools by litigants or judges can have substantial effects on settlement outcomes and *ex ante* behavior.

We now consider some extensions of these concepts and discuss further considerations upon introducing additional potential complexities into the model.

4.1 Factual Determinations

We have until now avoided applying our model to fact finding. But surely some algorithms might predict how a judge or jury would rule on certain facts in light of the available evidence and other characteristics of the case. The prediction here is more complicated and requires a more advanced algorithm with more data because it must take into account jury composition as well as the many small variations possible in the presentation of evidence.[34] But there may be some cases where a judge is the factfinder and the available evidence is similar across a large body of law. For example, some small claims cases may have these characteristics. In any event, if algorithms could predict the outcome of fact finding, the analysis would be similar to what we have presented.

To see why, assume a case where the law is well established and the parties have stipulated to all but one fact. In that case, the decision on that one fact will determine the outcome of the case. If the

factfinder rules one way, the plaintiff wins. If the factfinder rules the other way, the defendant wins. The parties will base their settlement strategy on their prediction of how the factfinder will rule on the fact in question. This scenario is essentially equivalent to that considered above, and the same analysis would apply.[35]

4.2 Variation and Uncertainty

When p^I is less than 100%, the outcome is uncertain. What does that uncertainty represent? Judicial bias or inconsistency? Missing variables in the analysis? Random human variation? Or meaningful human insight that data fails to capture? The answer to those questions may change litigation outcomes. But it may not matter in an environment where all cases settle. A settling party might not care about theoretical explanations for variation. If the variable cannot be measured, it is simply considered litigation risk and will be priced into settlement.

4.3 Judge-Specific Variation and Other "Excludable" Factors

Courts may reach different outcomes across similar cases for various reasons, as discussed above. This may be a function of inconsistency or it may simply be based on factors that are not represented in the data. One factor that likely will be identified is the judge's identity. We have treated courts as a monolithic institution throughout legal history. However, each judge is an individual. One judge might treat an identical case differently than another. To the extent the data can identify the outcome probabilities attributable to the identity of the judge, this information poses interesting questions.

First, it certainly complicates option 6 to ask how a judge would react to and implement p^I if she knows that the inputs of p^I include her own idiosyncratic preferences.

Second, we might want to design the algorithm – to the extent possible – to exclude or control for judge-specific effects. If outcomes are influenced by the mere identity of a particular judge, we may consider de-biasing results, correcting for that influence.[36] It is worth noting that doing so would affect some of the results in our models. Litigants using algorithms to predict success today, of course, do not want to exclude judge-specific effects. To the contrary, they are quite interested in the preferences of the judge to whom the case has been assigned.[37] Thus, current real world settlement estimates include and may be primarily driven by judge-specific effects. Removing those effects from judgments will produce results distinct from the baseline models. Again, option 6 becomes very complicated if the parties in settlement negotiations are predicting judge-specific reactions versus a model specifically corrected against judge-specific effects.

Similarly, the algorithms may reveal other undesirable factors that have been deciding outcomes of past cases. For example, prior outcomes may have been driven by improper factors such as race, gender, or the identity of one party's lawyer, or by trivial factors such as the date of filing.[38] To the

extent these factors can be identified and corrected for, using algorithmic assessment tools can add value to the judicial process.

4.4 Accuracy: Selection Bias

Accuracy is a critical element affecting the utility of these algorithms and we have avoided this consideration until now. What if the litigation prediction tool is not an accurate reflection of the way a court would have resolved the dispute? That is, what if the past decisions are not predictive of the future judgments? A litigation prediction tool only analyzes cases that *resulted* in a judgment. This leads to issues of selection bias. Perhaps only cases with edge-cases or confounding factual situations end up in court; straightforward cases may not. Thus, the predicted probability may generate accurate outcomes for cases actually going to court, but those probabilities may not reflect the entire landscape of legal disputes. This affects settlement outcomes. Lawyers and judges who use these tools therefore still require an understanding of the case law upon which these predictions are made. If the case at hand is different from all the prior cases, or if there are new facts that haven't been addressed before by the judiciary, this will weaken the predictive value of the algorithm. This is, of course, true of any predictions based on precedent, irrespective of whether data analytics are used.

4.5 Accuracy: The Choice of Statistical Model

We have assumed the existence of one objective probability, p^I, that is produced by the algorithmic assessment tool. In reality, there is no single p^I. Moreover, different choices of statistical models in the algorithm can change the probability output. This in turn will change the outcome of cases under the various models discussed. The effect is likely to be the greatest in option 2. The dollar value of a suit in option 2 directly depends on the independently assessed probability at all levels of p^I, not just around the neighborhood of $p^I = 50\%$. To understand the importance of the statistical model selection, consider the choice between a *probit* or logistic regression model. The imputed probabilities that are returned by using *probit* or logistic regression models often return similar probabilities, especially in the neighborhoods of 0%, 50%, and 100%, but they are not the same. For some fact patterns, logistic regression classification models will present a higher predicted probability than a *probit* model. Which model should be used here? More advanced machine learning classification models will create more confounding questions on probabilities. And these probabilities will vary with regularization and tuning. The choice of one model or the other to generate probabilities would favor one side. Simply put, with every decision in statistical modeling, there will be winners and losers.

4.6 Legal Rigidity and Stale Precedent

If litigation prediction tools lead more cases to settle or if judges use them to decide cases, judges will produce fewer decisions with precedential value. This may be costly. Case law may benefit from being dynamic and frequently updated.[39] And litigation prediction tools may impede those updates, leading

to law becoming stale. This could have two effects. The case law may no longer fit the world it governs. A common example is the development of traffic rules, initially geared toward the horse and buggy age and now ill-equipped to manage the age of automobiles. This problem would grow as judges begin adopting the litigation prediction tools, themselves. Alternatively, judges may ignore older, stale precedent (and the litigation prediction tools based on that precedent). If that happens, the litigation prediction tools – which use precedent as their input – will become less accurate, and parties will have new incentives to litigate. This could lead to a circular effect: prediction tools may at first be very accurate and highly utilized, depreciate as precedent becomes stale, go out of use, but then reacquire usefulness as precedent is updated.

4.7 Fee Shifting: Another Potential Use

Rather than using the prediction tool to make determinations about outcomes, judges may use them to make determinations about awarding costs. The judge may use the *ex ante* independent assessment of the plaintiff's likelihood of victory to award costs to the ultimate victor. If the algorithm suggests that a losing plaintiff had a very low likelihood of victory all along, the judge may elect to award costs to the defendant.

5. Conclusion

Algorithms don't make decisions. Rather, humans can make decisions that take algorithmic assessments into account. The way in which judges use algorithmic assessments of liability is not a simple "yes or no" decision. There are different considerations that judges must examine in order to effectively convert algorithmic predictions into legal decisions. We have endeavored to explore some of the possible choices available to the legal system.

The simple model presented here reveals that the use of algorithmic assessments of liability and the advent of automated judging, will have complicated and dynamic effects on settlement practices and litigation outcomes. In turn, these effects will alter *ex ante* incentive dynamics and deterrence. Further models, allowing for endogenous costs, sequential bargaining, asymmetric access to technology, and hidden information, will no doubt complicate these effects further. We view this model as an important introductory framework for exploring and understanding this new technology.

Appendix: Settlement where Litigants – But Not judges—Use Algorithmic Liability Assessments

Parties Use Litigation Prediction Tools[40]

Suppose the parties have access to litigation prediction tools, but judges and other decision makers do not. Let us also suppose that use of such tools by litigants is free of cost. The tool provides an

independent assessment of the likelihood of a plaintiff succeeding, p^I. Before stage 1, both parties update their prior subjective probabilities in light of this new independent assessment:

$$p_i^P = (1-\theta)p^P + \theta p^I$$
$$p_i^D = (1-\tau)p^D + \tau p^I$$

The degree to which parties adjust their assessments, θ or τ, depends on the degree to which each believes that the independent assessment accurately reflects the law. If parties are highly influenced by the independent assessment, then θ and τ are close to 1. If parties are skeptical of the algorithm, and trust their own intuitions, θ and τ are close to 0.

Typical Scenario – Likelihood of Settlement Increases

The effect of this adjustment depends on the relative position of p^I compared to the two prior subjective assessments of the parties. Initially, we assume that the independent probability assessment of the plaintiff's chance is at least as great as the defendant's subjective view; at most, the probability is as high as the plaintiff's subjective view:

$$p^I \in \left[p^D,\ \mathrm{p}^P\right]$$

We call this the "typical" scenario, as at least one of the parties is likely overly optimistic in their subjective assessment of the law. Under these assumptions, the likelihood of settlement increases. Recall that settlement is more likely to fail when the area of disagreement between the parties, Δp, is large. Both posterior probabilities of the plaintiff and defendant (weakly) converge towards p^I. The defendant's posterior probability is greater than her previous assessment while the plaintiff's posterior probability is lower. The new area of disagreement after the probability adjustment, Δp_i, is no greater than before.

It is trivial to show that when at least one of θ or τ are greater than zero, using litigation prediction tools will result in additional case settlements than would otherwise have occurred. The area of disagreement shrinks while the cost-damage ratio ($\frac{c^P+c^D}{L}$) remains constant. This results in more cases being within the conditions of equation (1). We show this graphically in **Figure 9**. Conversely, there are no disputes that would have settled in a world without litigation prediction tools but do not settle once such tools are introduced.

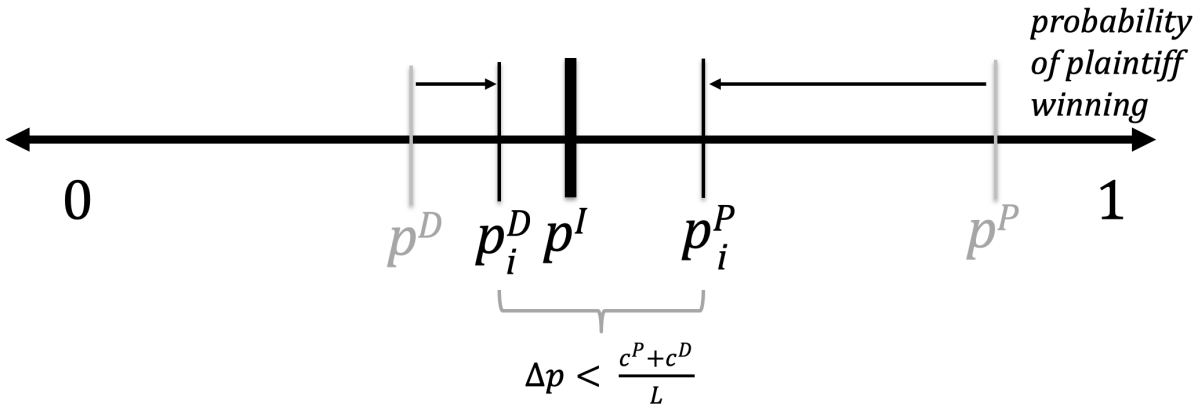


Figure 9: The independent assessment of the plaintiff's probability of victory is p^I. The defendant updates her probability assessment up from p^D to $p^D{}_i$. The plaintiff revises her subjective probability analysis down from p^P to $p^P{}_i$. The area of disagreement decreases. Here, if the new area of disagreement is smaller than the cost-damage ratio, the two parties settle at equilibrium.

At equilibrium, if settlement is possible, the defendant makes a "take-it-or-leave-it" offer in stage 2 of S_i^*:

$$S_i^* = \left((1-\theta)p^P + \theta p^I\right).L - c^P$$

The plaintiff accepts this offer. The equilibrium settlement amount increases in p^I, with a slope of θ. If the plaintiff disregards the new information and does not update her position ($\theta = 0$), then the equilibrium settlement is constant at $S_i^* = p^P.L - c^P$. But if the plaintiff treats the independent assessment as definitive and updates their position to conform to the independent analysis ($\theta = 1$), the settlement offer is $S_i^* = p^I.L - c^P$. This latter case is shown in **Figure 10.** These settlement equilibria are conditional on settlement being possible (i.e., the condition in equation (1) is satisfied).

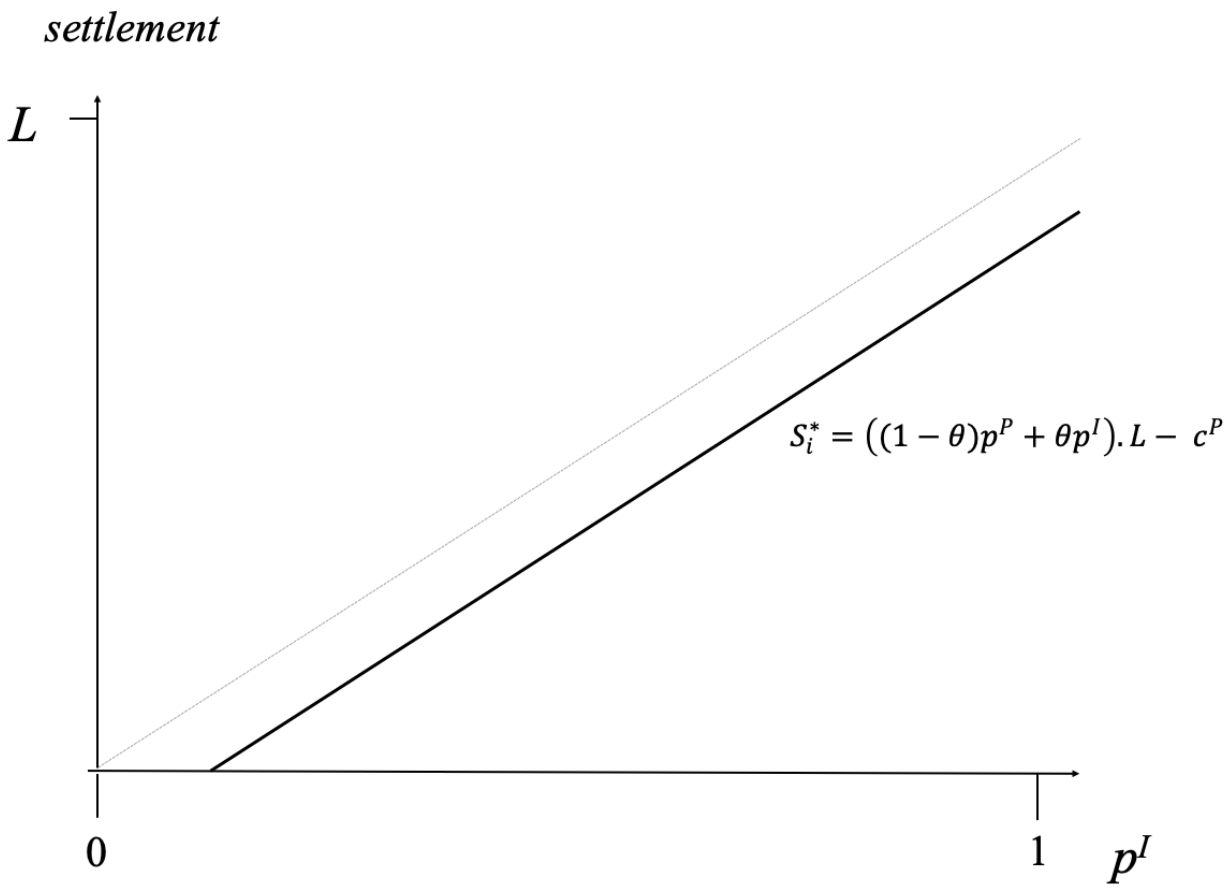


Figure 10: Settlement offers when $\theta = 1$. The settlement offers track p^I.

Atypical Scenario—Likelihood of Settlement Decreases

There are situations where prediction tools reduce the likelihood of settlement. Here, we relax the assumption that the independent assessment falls between the two subjective prior assessments. Take, for example, a situation where the defendant's prior evaluation reflects a relatively pessimistic view (high p^D) and the independent assessment is even lower than the defendant's:

$$p^I < p^D$$

There are now situations in which the parties would have settled without litigation prediction tools, but no longer will. To understand why this change would occur, assume that the defendant's pessimistic prior is the same as the plaintiff's optimistic prior, such that:

$$p^D = p^P$$

Without litigation prediction tools, a settlement would be certain because the area of disagreement is zero and equation (1) would be satisfied in all cases. But the introduction of the independent assessment p^I has the potential to upset this equilibrium. Let us stipulate $p^I < p^D$. If the defendant

updates their position with high conformity to the independent assessment ($\tau$is high) and the plaintiff does so weakly ($\theta$is low), then after updating:

$$p^I < p_i^D < p_i^P$$

The area of disagreement increases. More generally, for any set of p^P and p^D, the area of disagreement will increase as long as $p^I < p^D$ and τ is greater than θ.

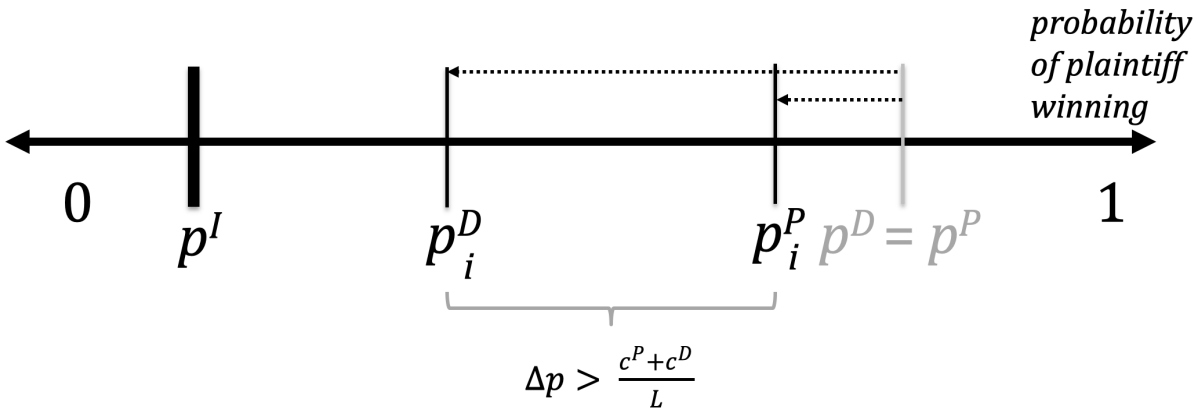


Figure 11: Initially, the two parties have the same subjective belief, $p^D = p^P$. There is no area of disagreement. With no access to litigation prediction tools, the dispute will settle. The litigation prediction tool, however, predicts that the probability of a victory for the plaintiff victory is much lower, p^I. The two parties update their prior assessments. This creates an area of disagreement. In this illustration, the new area of disagreement is sufficiently large to prevent settlement from occurring.

Similarly, settlement opportunities will be reduced when the plaintiff is pessimistic relative to the independent probability assessment ($p^P < p^I$), and she updates her prior assessment with strong conformity to the independent assessment while the defendant updates with weak conformity.

The atypical scenario relies on the condition that the pessimistic party (defined as the party whose initial belief in her own success is worse than the independent assessment) updates their assessment with stronger conformity than the optimistic party (whose initial belief in her own success is better than the independent assessment). Readers may question the likelihood of this condition. But there are good reasons, from a behavioral perspective, to think that Bayesian updating may be asymmetric. Optimism bias on the part of the parties can produce these results. Parties may update their prior assessments more weakly in the face of bad news than in the face of good news.[41] For example, plaintiffs may be more willing to gravitate toward higher independent assessments of success versus lower assessments. If this is true, then *provided that* the independent assessment falls outside the bounds of the two subjective assessments, litigation prediction tools may reduce the opportunity for settlement. Our point here is not that settlement will always fall—it is merely to show that there are atypical situations where the opportunity for settlement falls.

The Number of Disputes Brought

There may be cases that would have been brought but for the litigation prediction tools (resulting in fewer disputes). The plaintiff files suit in this model when her expected return is greater than zero:

$$p^P.L - \ c^P \geq 0$$

In the typical scenario, we would expect that the litigation tool would temper the plaintiff's optimistic prior assessment, reducing the likelihood of bringing suit. But, as demonstrated above, there may be situations in which the converse is true. If the plaintiff was sufficiently pessimistic without litigation prediction tools, the use of such tools may actually encourage plaintiffs who would not have otherwise brought a claim. Litigation prediction tools may even be used to discover potentially successful cases of which the plaintiff was not aware (imagine for example a plaintiff who did not realize a certain tort was legally actionable). Essentially, the litigation tool serves to inform the plaintiff of her legal rights.

To the extent that the independent assessment is correlated with the actual outcome, these effects should enhance the welfare of claimants. Disputes where the plaintiff has a weak claim (low p^I) are less likely to be brought, while disputes where the plaintiff has a strong claim (high p^I) are now more likely to be brought.

Footnotes

1. *See, e.g.,* Christopher Markou, *Are We Ready for Robot Judges?*, Discover Magazine (May 15, 2017) available at https://www.discovermagazine.com/technology/are-we-ready-for-robot-judges; Larry Mantle, *Can a Robot Make a Fair Verdict?* Airtalk Podcast available at https://www.scpr.org/programs/airtalk/2019/04/01/64335/can-a-robot-judge-make-a-fair-verdict/. ↩

2. Eric Niler, *Can AI Be a Fair Judge in Courts? Estonia Thinks So*, Wired (March 25, 2019) available at https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/; Victor Tangermann, *Estonia is Building A "Robot Judge" to Help Clear Legal Backlog*, Futurism.com (March 25, 2019) available at https://futurism.com/the-byte/estonia-robot-judge. ↩

3. Monisha Pillai, *China Now Has AI-Powered Judges*, RADII (Aug. 16, 2019) available at https://radiichina.com/china-now-has-ai-powered-robot-judges/; Chris Young, *China has Unveiled an AI Judge that Will 'Help' With Court Proceedings*, Interesting Engineering (Aug 19, 2019) available at https://interestingengineering.com/china-has-unveiled-an-ai-judge-that-will-help-with-court-proceeding; see also Jingting Deng, *Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System*, 3 Geo. L. Tech. 223 (2019); Tom Fish, *AI shock: China Unveils 'Cyber Court' Complete with AI Judges and Verdicts via Chat App,* Express.com (Dec. 6,

2019) available at https://www.express.co.uk/news/science/1214019/ai-china-cyber-court-artificial-intelligence-judges-verdicts-chat-app. ↩

4. Henriette Nakad-Westrate, Ton Jongbloed, Jaap van den Herik, Abdel-Badeeh M. Salem, *Digitally Produced Judgements in Modern Court Proceedings*, 6 Intn'l J. of Dig. Soc. 1102 (2015). ↩

5. *See generally* Tania Sourdin, *Judge v Robot? Artificial Intelligence and Judicial Decision-Making*, 41 UNSW L. J. 1114 (2018); *Council of Europe Adopts first European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems*, Council of Europe (Sept. 13, 2019) available at https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. ↩

6. *See, e.g.*, Dru Stevenson & Nicholas Wagoner, *Bargaining in the Shadow of Big Data*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325137 (2014); Ben Alarie, Anthony Niblett, & Albert Yoon, 58(3) Canadian Business Law Journal 231 (2016); Sourdin, *supra* note 5 at 1125. ↩

7. Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 Ga. St. U. L. Rev. 1305, 1307 (2019), Rebecca Crootoff, *"Cyborg Justice" and the Risk of Technological-Legal Lock-In*, 119 Colum. L. Rev. Forum 233 (2019). For a general discussion of law firm use, see Bernard Marr, *How AI and Machine Learning Are Transforming Law Firms and the Legal Sector*, FORBES, available at https://www.forbes.com/sites/bernardmarr/2018/05/23/how-ai-and-machine-learning-are-transforming-law-firms-and-the-legal-sector/#6308a31c32c3. ↩

8. Jingting Deng, *Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System*, 3 Geo. L. Tech. 223 (2019); Eugene Volokh, *Chief Justice Robots*, 68 Duke L. J. 1134 (2019); Crootoff, *supra* note 7. Separately, judges have also started using algorithms that provide information about other dimensions of litigation. See Megan Stevenson, *Assessing Risk Assessment in Action*, 103 Minn. L. Rev. 2018); Alex Albright, *If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions,* (2019) available at https://thelittledataset.com/about_files/albright_judge_score.pdf; see also Rachel Traughber, *Finding a Link to the Human in Algorithms Setting Justice*, (2018) available at https://news.harvard.edu/gazette/story/2018/05/grad-discovers-algorithms-in-justice-system-dont-always-compute/. ↩

9. Courts in China are reported to have begun using smart "intelligent case-deciding programs" to provide guidance to judges to ensure that similar cases are decided in the same manner. Deng, *supra* note 8. ↩

10. Deng, *supra* note 8;Volokh, *supra* note 8; Richard Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 Stan. Tech. L. Rev. 242 (2019); Crootoff, *supra* note 7. Estonia has also

proposed the use of automated "robot judges" in small private disputes, although it is not obvious whether the system will be based on prior case data or simply a coded decision algorithm. ↩

11. We will refer to "judges," "courts," and judicial decisions throughout this paper. The results of our analysis apply similarly to other arbiters resolving two-party legal disputes. Thus, administrative judges and other regulatory tribunals as well as arbitrators in binding arbitration would fall under the analysis. ↩

12. To be sure, there are other tools that might assist judges in reaching a decision. Two in particular come to mind. First, for simple cases, structured code might simply assess whether certain defined factual elements are present and output a binary judgment. This is certainly possible for very simple cases. (For example, was the car parked in a prohibited spot? If yes, then "liable.") But for cases involving more variables, the task of creating such code becomes more difficult. Writing structured code is an obstacle for any cases that turn on behavioral standards rather than explicit rules. (For example, was the conduct reasonable?). Second, in theory an advanced machine learning algorithm could be programmed to achieve certain objective and adjudicate cases based on that objective. Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, 86 U. Chi. L. Rev. 333 (2019); Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 Ind. L. J. 1401 (2017). Such technology does not currently exist in any form that could be practically implemented. We have focused on the predictive algorithms based on prior case data because this technology is currently in use by lawyers and even by judges in some foreign jurisdictions. ↩

13. The same analysis can be applied to fact finding. But such analysis—and likely the algorithms—are more complicated, and so we bracket that application for now. We will return to the fact finding application in Section 4. ↩

14. *See generally* Andrew F. Daughety & Jennifer F. Reinganum, *Settlement*, *in* 8 Encyclopedia of Law and Economics 386, 386–71 (Chris W. Sanchirico ed., 2nd ed. 2012). While the model we present is simplified, it is intended to capture the essential elements of uncertainty, optimism, and asymmetric information from prior literature. *See, e.g.*, Robert Mnookin & Robert Wilson, *A Model of Efficient Discovery*, 25 Games & Econ. Behav. 219, 220 (1998); Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs*, 11 J. Legal Stud. 55, 63 (1982); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. Legal Stud. 399, 422 (1973); J. J. Prescott, Kathryn Spier, Albert Yoon, *Trial and Settlement: A Study of High-Low Agreements*, 57 J.L. & Econ. 699 (2014); Abraham L. Wickelgren, *Law and Economics of Settlement*, *in* Research Handbook on the Economics of Torts 331–32 (Jennifer Arlen ed, 2013). ↩

15. We discuss applications to fact finding below in Section 4. ↩

16. We do not invest significant time assessing the quality of the argument nor particular facts of the case affecting argument credibility. ↩

17. *See generally* Daughety & Reinganum, *supra* note 14. ↩

18. The function here is an obvious oversimplification. However, it captures the relevant elements of the problem that will be necessary later to demonstrate the effects of judicial use of predictive algorithms. ↩

19. At this stage of analysis, this assumption is relatively inconsequential. ↩

20. This is a common type of game theoretic model in the law and economics literature. *See e.g.*, Lucian A. Bebchuk, *Litigation and Settlement Under Imperfect Information*, 15 RAND J. Econ. 404 (1984); I. P. L. P'ng, *Behavior in Suit, Settlement, and Trial*, 14 Bell J. Econ. 539, 541 (1983); Jennifer F. Reinganum & Louis L. Wilde, *Settlement, Litigation, and the Allocation of Litigation Costs*, 17 RAND J. Econ. 557 (1986). ↩

21. *See* Bebchuk, *supra* note 20, at 407; Alan E. Friedman, *An Analysis of Settlement*, 22 Stanford L. Rev. 67, 92 (1969); P'ng, *supra* note 4, at 544; George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. Legal Stud. 1, 13 (1984); Shavell, *supra* note 14, at 63; William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & Econ. 61, 66 (1971); John P. Gould, *The Economics of Legal Conflicts*, 2 J. Legal Stud. 279, 284–86 (1973); Posner, *supra* note 14, at 419; Joel Waldfogel, *Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation*, 41 J. Legal Stud. 451, 455 (1998). ↩

22. We do not address the scenario where judges use the tools, but parties do not. The model in such a situation would be similar to the baseline model with subjective probabilities. But the parties would no longer be predicting simply what the judge will do, but rather what the judge utilizing the tool in a certain way will do. The size of settlements remain the same, but the parties will settle at different rates because the prediction of a judge's behavior differs when the judge has access to algorithmic tools versus when she does not. ↩

23. This statement is based on anecdotal conversations. We are currently devising a survey which more rigorously tests people's intuition. ↩

24. Moreover, even if we were modeling fact finding, the prediction tool would not provide a likelihood that a particular fact was true. In a generic civil case, a prediction that a court is 51% likely to find a certain fact to be "true" is akin to the contention: "it is more likely than not that a court will think that a fact is more likely than not to be true." This contention amounts to a drastic change in the burden of proof. The added layer of probability makes interpretation more complicated. For

example, what if there was one case with a 40% probability that a court will find a certain fact has an 80% probability of being true versus another case with an 80% probability that a court will find a certain fact has a 40% probability of being true? It is hard to say which is a normatively stronger case. But from a litigating plaintiff's point of view, the first is much stronger than the second because it has at least a 40% chance of winning whereas the second has at least an 80% chance of losing. We discuss the fact finding application briefly in Part 4. ↩

25. The proof of this is relatively straightforward. For any given p^I, where p^I is less than 50%, the fraction of cases where the results are different = p^I. For any given p^I, where p^I is greater than 50%, the fraction of cases where the results are different = 1 – p^I. The total fraction of cases is found by taking the integral of this inverted-v curve. This is equal to ¼.

Of course, one might consider alternative distributions. For example, the distribution might instead have a barbell formation with cases clustering around 1% and 99%. This would suggest that most cases are easily adjudicated. If that were true, the number of cases decided differently, and the magnitude of the difference in most cases, would decline. ↩

26. Or let themselves drift toward the 100% mark. ↩

27. For an in-depth discussion of the potential role of randomness in law more generally, see Neil Duxbury, *Random Justice: On Lotteries and Legal Decision-Making*, (Oxford University Press 1999). ↩

28. Some "variation" in outcomes might be explained by judge-specific effects. We do not discuss those effects here. But we return to them – and to the question of whether entrenchment of judge-specific effects in such an algorithm is socially desirable – in Part 4. ↩

29. This is the state of real-world litigation today. Most cases do settle. And they settle based on the parties' predictions about the outcome, which take into account judicial variation regardless of its source. ↩

30. The magnitude of the reduction is based upon the distribution of easy cases. Triage is particularly effective if most cases are easy, creating a distribution with a barbell formation. ↩

31. For discussion on how settlement bargaining affects incentives to bring frivolous suits, *see* Avery Katz, *The Effect of Frivolous Lawsuits on the Settlement of Litigation*, 10 Int'l. Rev. L. & Econ, 3–27 (1990). ↩

32. *See* Appendix. ↩

33. *See* Daughety & Reinganum, *supra* note 14. ↩

34. Parties may also expend resources to develop new evidence to influence the algorithms. But the same can be said of expenditures to develop new legal arguments. ↩

35. Note that the burden of proof does not affect the analysis here. It may change the probability of a judicial ruling but does not affect the strategic position of the parties. ↩

36. For a discussion of related concepts, see Daniel Chen, *Machine Learning and the Rule of Law*, (2019) available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3302507&download=yes. ↩

37. The question of predicting idiosyncratic judge effects is controversial in some jurisdictions. In France, a recent law prohibited any parties from using predictive litigation algorithms that take into account the individual identities of judges. Jason Tashea, *France Bans Publishing of Judicial Analytics and Prompts Criminal Penalty*, ABA Journal, June 7, 2019 at https://www.abajournal.com/news/article/france-bans-and-creates-criminal-penalty-for-judicial-analytics. ↩

38. *See, e.g.*, Anthony Niblett, *Algorithms as Legal Decisions: Gender Gaps and Employment Law in the 21st Century,* Univ. of N. Bruns. L. J. (2020, forthcoming). ↩

39. *See, e.g.*, Shavell, *supra* note 14 on the social benefits of trials and setting precedent. ↩

40. The analysis in this Appendix is generally consistent with Stevenson & Wagoner, *supra* note 6 and other settlement models. It is presented here as a comparison to our new analysis of the judicial use of algorithms in this paper. ↩

41. Many studies have shown that people tend to exhibit confirmation bias by updating their prior assessments of likelihood more frequently when receiving new evidence consistent with a preferred outcome. *See e.g.*, Tali Sharot et al., *How Unrealistic Optimism is Maintained in the Face of Reality*, 14 Nature Neuroscience 1475 (2011); Matthew Rabin & Joel L. Schrag, *First Impressions Matter: A Model of Confirmatory Bias*, 114 Q. J. Econ. 37, 37–38 (1999); David Eil & Justin M. Rao, *The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself*, 3 Am. Econ. J. Microeconomics 114 (2011). ↩